

Getting to grips with fraud

Dealing with scientific fraud is likely to be an unpleasant process for accused and accuser alike. The only solution lies in well conceived methods of procedure.

UNTIL recently there would seem to have been little need for discussion of the dangers of the fraudulent reporting of data to the scientific enterprise. There was always the odd historical case of fraud and every generation seemed to bring forth at least a handful of people who were determined to identify finally the perpetrators of the Piltdown Man hoax. But that was a pastime conducted in a similar obsessive spirit to the hunt for the identity of the "dark lady" of Shakespeare's sonnets and seemed to have little relevance to the daily practice of the scientific profession.

In retrospect, 1980 proved a turning point with allegations of falsification of data made in four major cases. The pace of discovery of new cases of fraud, malpractice and plagiarism has not slackened subsequently. Since 1980, such allegations have been made publicly in at least twenty-five cases and cleared up privately in a further seven cases.

These are the figures, not necessarily comprehensive, that obtained at a workshop on scientific fraud and misconduct sponsored jointly by the American Association for the Advancement of Science and the American Bar Association last week. The figures are not by themselves particularly worrying. Even though more than a hundred papers tainted by fraud may have been produced, that number is still insignificant compared to the world output of scientific literature. But there remain several issues to worry about.

How are scientific frauds detected? Conventional wisdom has it that science is self-correcting: experiments are repeatable and a fraudulent or incorrect result will inevitably be detected in a later repetition. But in much of the biological sciences, where a great measure of skill and experience is needed to master new techniques, things are not so simple. And in the medical sciences, where human subjects are involved, experimental opportunities may have to be taken as they arise, and cannot be totally controlled or perfectly repeated. To say these things does not call into question the basic principle that experiments are repeatable. But it does mean that the first thought of someone who fails to replicate a result is not likely to be that the result is fraudulent. A more charitable explanation is likely to be sought in minor differences in experimental conditions.

For these reasons it appears that frauds have rather rarely been discovered through the replication of experiments by totally independent groups. The discovery of fraud is more usually a personal affair: a co-worker's suspicions are aroused, and then confirmed by events extraneous to the published record — for example, that a reagent essential to the experiment had not been ordered at that time, or even that the number of Petri dishes used in an experiment was greater than the number the laboratory had in stock.

The personal nature of the discovery of fraud inevitably makes it a painful business. It is not like having been a chance witness at a bank robbery. No one is going to emerge as a hero from an investigation which is likely to blacken not just the name of the individual scientist but also his or her institute and the scientific enterprise itself. But that should not mean that the person who exposes a fraud is penalized for his or her courage.

Unfortunately, the experiences of "whistle-blowers", personally related at the workshop, makes clear that the price of

exposing fraud is high. Action by institutions, which clearly saw their own reputations at stake, was often taken grudgingly and after years of delay. The motives of the accuser were frequently questioned: did jealousy or personal animosity play a part?

There seems no escaping the conclusion that for both accuser and accused to have full protection of their good reputations, which is ultimately all that a scientist has to assure continued employment, universities need effective processes for dealing with allegations of fraud and dispensing justice rapidly. Some universities — mainly those who have been through the protracted trauma of a fraud case — have already made a start but their experiences show that it is not easy. The skills for investigating fraud, different from those necessary for research, are hard to obtain. And universities have genuine conflicts of interest between dispensing justice and protecting their reputation that may make systematic policies and procedures unpalatable.

There is little hope that universities transfer their responsibility to investigate fraud to other bodies (except when the misuse of public funds makes it necessary to involve the granting agencies). Scientific societies may help to set guidelines of proper practice but they are not in a position to deal with individual cases. Nor can a great deal of help be expected from scientific journals. Although journals may choose to publish letters that point out inconsistencies in, or failures to replicate, previously published work, they would find themselves in a difficult legal position if they began to make accusations about their own authors.

Could journals do more to ensure that fraudulent work never appears in print? "Why ever was the paper accepted" is a natural enough reaction when a case of fraud is uncovered. But evidence of fraud is not necessarily to be found in the published account. And while conscientiousness can be expected of referees, the total scepticism needed to suspect that an author's every statement may be false, surely cannot. Even scepticism may be insufficient for fraudulent work is often correct in its conclusions. That need cause no surprise. People who commit fraud are only very rarely stupid enough to falsify scientific breakthroughs. More characteristic is a pedestrian piece of work with a wholly expected result. But that that is so forces a hard question: How different is fraud from shoddy practice?

If the timid manufacture of data to conform to already probable hypotheses is the most common form of fraud is it one end of a spectrum of shoddy practice that includes the selection of data, the filling in of inadequate or control data through phantom experiments, and the use of guest co-authors to add credibility to mediocre work? The more comforting view is that frauds are committed by a few highly deviant individuals whose activities have nothing to do with the normal scientist. Between these two hypotheses — that fraud represents either the tip of an iceberg or a few bad apples — there are not data to make a judgement, although practising scientists clearly each have their own view. But a corollary of the first hypothesis is that there are rewards for the mediocre work that no one will ever read carefully. That leads to a much larger question: has the scientific enterprise come to stress quantity to such an extent that quality has begun to suffer? □



Japanese industry dreams of superconductor future

■ Parliament lobbied for more cash

■ Flood of pre-emptive patents

Tokyo

How does Japan foresee the development of high-temperature superconductors? A recent meeting between scientists and members of the House of Councillors provides some clues.

At the end of last month, four leading superconductor researchers briefed members of the special committee for science and technology of the upper house of the Diet. Ruichi Nakagawa, Director of the Science and Technology Agency's National Research Institute for Metals, began the 3½-hour session with a simple explanation of superconductivity. To convey the importance of zero resistance to the Diet members he pointed out that thermal losses due to resistance in Japan's national electricity grid equal half the electricity consumption of Tokyo, or ¥600,000 million a year (more than \$4,000 million). Professor Kazuoki Fueki of Tokyo University explained resistivity at the atomic level and the notoriously complex 'BSC theory' of superconductivity, while Michihiko Nagumo of Shin Nippon Steel and Shigeru Hayakawa of Matsushita Electric Co. pointed out the potential practical applications.

The four researchers were unanimous in stressing the need for more basic research into the properties of the new materials and a theory to explain high- T_c superconductivity. They also called for international collaboration in research. The barriers to development of practical applications, such as the low critical current densities of the bulk oxides, their fragility and sensitivity to oxygen and the difficulties of processing them into wires, were emphasized. But this did not stop Hayakawa "dreaming" about the future. He foresees applications including gigantic liquid-crystal displays networked with superconducting wire, high-density memory devices such as centimetre-long video tapes with superconducting memory loops, supersensitive Josephson junction devices capable of monitoring the magnetic fields of the brain and heart, high-speed Josephson computers, and high-definition satellite television.

Professor Fueki complained that lack of money is the major problem besetting basic research in the universities, a view echoed by Nakagawa, who pointed out that high-field magnets to measure magnetic-field properties of the new materials cost several hundred million yen (several million dollars) each. And the Science and

Technology Agency has recently applied for over ¥1,000 million to build three high-field magnets at Nakagawa's institute next year (see *Nature* 329, 93; 1987).

On the issue of patents, there were fears that the United States may rush to develop technology and shut out Japan. Nagumo argued that the new superconductors could lead to a new industrial revolution, and that patents should not be allowed to block the way; instead there should be international "give-and-take". Hayakawa called for "open patents" that allow licensing by anyone, and he suggested that rapid development of consumer products could forestall the classification of superconductor applications as military secrets. But Sumitomo Electric Co., a cable manufacturer, is taking no chances. The company has already filed over 600 patent applications for the new superconductors



in the hope of blocking US patents and arranging cross-licensing agreements.

Japan also fears that it may prove difficult to secure adequate supplies of the raw materials to make high- T_c superconductors. In reply to a question from house member Masao Goto, Nagumo said that Japan will need to import several million tons per year of rare earth elements such as yttrium, as opposed to the few thousand tons currently imported.

The meeting closed with a question the experts could not answer. Hiramatsu Deguchi asked about the effects of magnetic fields on people's health. Both Fueki and Nakagawa admitted that the medical effects of magnetic fields are unknown, and they called for more research into this matter.

David Swinbanks

Academy pushes basics too

Washington

US government agencies are in the midst of a debate over how best to pursue the potential of the newly discovered class of superconducting materials. The National Academy of Sciences has now weighed in with its perspective*, urging the government not to ignore the science of superconductors while it pursues their technological promise. The Academy report also cautions that interest in superconductors could place demands on the scientific and technical manpower pool beyond its capacity to respond.

The report, produced by a panel of the committee on science, engineering and public policy, says a national programme to exploit the new materials should focus on a better understanding of their essential properties, as well as the mechanism responsible for those properties. In addition to searching for new superconducting materials, the report suggests starting the process of creating prototype devices that might incorporate the new materials. For the next fiscal year, the panel reckons that \$100 million will provide "a good beginning" in exploring superconducting compounds and their potential uses.

The rallying cry of improving economic competitiveness has frequently been sounded to justify research on superconductivity. Both Congress and the Reagan administration are anxious that research advances be translated into commercial products in a timely fashion. But the Academy report points out that the world superconductivity market is relatively small, although products made with superconductors — especially medical imaging devices — represent a much larger market. Given the engineering problems presented by the ceramic superconducting materials, the panel expects it will take a decade of precommercial exploration to establish the range of potential applications. Although the panel is unanimous that the new materials are bound to result in new economic growth, it will take room-temperature superconductivity to bring about a revolution.

The attraction of working with superconductors, both because of the intellectual excitement they have generated and the financial backing they are likely to attract, may syphon off engineers from other research areas where they cannot easily be replaced. The report urges that attention be paid to the educational demands that must be met if adequate numbers of suitably trained people are to continue to enter the field. **Joseph Palca**

*High-Temperature Superconductivity, Committee on Science, Engineering, and Public Policy, National Academy of Sciences, Washington, D.C.

Pakistani satellites

PAKISTAN is close to making a decision on plans for the launch of a pair of direct television broadcast satellites. According to the chairman of Pakistan's Space and Upper Atmosphere Research Commission (SUPARCO), Salim Mehamood, the Parliament will soon have to debate the funding share of the various agencies that will use the two "Paksat" satellites.

Pakistan has been stimulated by Indian successes with its SLV-3 rocket, which Pakistan claims could be used as an intermediate-range ballistic missile, to give more weight to space development. A ten-year plan stressing the need for indigenous launch capacity has been released and earlier this year a new facility for the production of solid fuel was inaugurated. At present, Pakistan's capabilities encompass only the production and launch of sounding rockets from its Sonamiani Beach base near Karachi. Paksat would be launched abroad.

Space development could well be hastened by cooperation with other Islamic countries. Serious efforts are being made towards the creation of an Islamic Space Institute with Bangladesh, Egypt, Saudi Arabia, Indonesia, Iran and Turkey as partners. R.R.

Collider sites

OF the 43 sites proposed for the Superconducting Super Collider (SSC), 36 have survived internal screening by the US Department of Energy (DoE), and have been passed on to the National Academies of Sciences and Engineering for more detailed assessment and reduction to a shortlist, to be announced next January (*Nature* 329, 92; 1987).

The only rejected site with official state backing was New York's proposal to build the SSC partly in Canada. Three New York locations wholly within the United States remain in contention. Bids were also disqualified if they did not fully demonstrate that the site could physically accommodate the SSC layout, had adequate electricity or water supply, would come free of cost, or was environmentally acceptable. Three of the unsuccessful bids were in Texas, but the state still has two official sites and two backed by county governments, as well as an interest in a site in New Mexico. D.L.

British space lobby

BRITISH industry is continuing to lobby the government for greater investment in space research. British Aerospace has written to 135 members of parliament and peers of the House of Lords pointing out the benefits to the economy of an expanded space programme. In particular, the need for Britain to at least retain its current position in the European Space Agency is emphasized. S.H.

Britain's genetic manipulation regulations to be extended

London

VOLUNTARY schemes for notification of the deliberate release or industrial use of genetically manipulated organisms in Britain are likely to be replaced by compulsory notification. Any failure to notify the Health and Safety Executive (HSE), whose commission proposed the changes last week, would become a legal offence.

The Health and Safety Commission, which put forward the changes after consultation with its Advisory Committee on Genetic Manipulation, says its main objectives are to ensure the proper risk-assessment of planned release and industrial use of altered organisms, and effective inspection of the establishments concerned.

But the changes are also designed to remove the anomaly that notification of the genetic manipulation of organisms is already compulsory, whereas the employment of the organisms, which may well carry a greater risk, is not. The changes are expected to anticipate the recommendations of the Royal Commission on Environmental Pollution that is due to report on deliberate release experiments within the next few months.

The HSE also hopes to make compulsory two other aspects of genetic manipulation that are currently voluntary. One would make it essential for anyone notifying the HSE of a genetic manipulation experiment to assess the risk of the experiment by a method approved by the HSE. The other would make it compulsory for any establishment that carries out genetic manipulation to set up a safety committee.

Other changes proposed in last week's consultative paper are designed to close two loopholes in existing regulations. No longer would it be possible, in theory, to argue that the direct incorporation (by, for example, microinjection) of foreign DNA into cells is not subject to regulation. And deliberate release from offshore installations (an oil rig, for example) would unambiguously be covered by the new regulations, although strangely they would not apply to ships in territorial waters or aircraft flying over them.

In one respect the new regulations would somewhat lighten the bureaucratic load on British genetic manipulators. Those involved in work that is categorized as low risk will need only to give numbers rather than details of their experiments in the annual retrospective notification they have to file.

Peter Newmark

■ The UK Department of Trade and Industry is hoping to interest companies in jointly funded experiments to assess the risk of deliberately releasing genetically manipulated plants or microbes. If experiments of sufficient interest to a number of

companies can be identified, the department's Biotechnology Unit is prepared to pay half of their cost.

Roy Dietz, head of the unit, anticipates that such experiments will be carried out under contract at independent laboratories, such as those of the Agricultural and Food Research Council or Natural Environment Research Council that already have experience in deliberate release experiments. A major aim of the "Planned Release of Selected and Manipulated Organisms" initiative, he says, is to hasten the evolution of a framework for deliberate release, which is being constructed on a case-by-case basis by the Advisory Group on Genetic Manipulation. □

Genentech lose Protropin case

Washington

A LEGAL dispute between Genentech and Eli Lilly over recombinant human growth hormone has prompted a review by Congress of the Orphan Drug Act, the law intended to give commercial protection to companies making treatments for rare diseases.

Genentech filed a case against the Food and Drug Administration (FDA) in March, after the FDA approved Eli Lilly's form of recombinant human growth hormone, Humatrope, under the Orphan Drug Act (see *Nature* 326, 231; 1987). Humatrope differs from Protropin by one amino acid, so it was considered a new drug by the FDA. But Humatrope is also used to treat dwarfism, and Genentech claimed that its marketing exclusivity for Protropin was being infringed.

Genentech last week learned that a complex legal strategy to protect Protropin from competition was a failure. Genentech sought to void both its own and Lilly's protection under the Orphan Drug Act, claiming that neither Protropin nor Humatrope was a "new drug", and that neither should be protected by the act. But a US Federal circuit court judge dismissed that argument, upholding the rights of Eli Lilly to market their drug.

Now the legislature is gearing up to clarify the Orphan Drug Act, to be introduced into the House on Tuesday. As *Nature* goes to press it is unclear whether the amendments will define the term "drug" in the context of the act, or establish a principle of "shared exclusivity", several drug manufacturers enjoying market protection for the same product for slightly different indications.

Carol Ezzell

French government shifts accent of spending towards industry

Paris

THE 1988 budget proposals presented last week by French minister for research and higher education, Jacques Valade, confirm the government's intention to shift the balance between private- and public-sector science and technology research in France. While several measures are envisaged to encourage industry-sponsored research, growth in state-supported research establishments (*grands organismes*) has been clearly restricted.

University teaching departments, which lack the infrastructure to carry out research, do slightly better in the budget. Grants for university-based research increases next year by 14 per cent, although this is less than the increase for 1987 (25 per cent).

The overall science and technology research budget will increase by 8.5 or 10.5 per cent compared to last year (depending on whether pledged sums are included in calculations), but this includes both civil and defence spending. When civil research alone is considered, the increase is a more modest 3.4 per cent, to FF39,345 million (\$6,503 million), against an annual inflation rate of about 2 per cent. Defence research spending, while less than in 1987, will still increase by 10 per cent to FF33,000 million.

Proposed grants to individual research establishments — still to be agreed by the Finance Ministry — vary. The Centre National de la Recherche Scientifique (CNRS), with its associated national institutes, and the Institut National de la Santé et de la Recherche Médicale (INSERM) — the two largest *grands organismes* — can expect increases of about 1 and 3 per cent, respectively. The Institut Français pour l'Exploitation de la Mer (IFREMER) will get an increase of about 2.6 per cent over 1987, while the nation's controversial technology 'shop window', the Cité des Sciences et de l'Industrie, at La Villette, Paris, will have its grant drastically cut by 11 per cent (see *Nature* 329, 96; 1987).

The lack of government favours for the *grands organismes* is also seen in the negligible number of new posts for young researchers for 1988 — 150 overall, against 398 in 1987. Speaking to the press, Valade considered the creation of new jobs "an achievement", since none had been expected, but forgot to mention the loss of 350 posts at engineer, technician and administrator grades.

Unlike the situation in most Western countries, researchers taken on by the *grands organismes* have jobs for life. Because of a high intake of new researchers in the 1960s and early 1980s, when public sector research had a higher government priority, present research teams are rela-

tively young and senior staff will retire at about the same time, in the late 1990s.

Unions representing researchers, as well as the independent advisory body, the Conseil Supérieur de la Recherche et de la Technologie (CSRT) argue that a much higher annual rate of job creation is necessary to offset the "massive" drop predicted for the end of the century. Valade, on the other hand, feels that "because growth in employment has been high it need not stay high" and wants to see posts liberated by mid-career researchers joining the private sector. Considerable financial incentives have been offered to researchers making such a move, and FF500 million of tax incentives are to be available to companies developing their research activity.

The government emphasis on commercially exploitable research is also apparent in a 50 per cent increase in grants to the Agence Nationale de Valorisation de la Recherche (ANVAR), set up to promote research and development in industry, particularly in association with public sector research groups. ANVAR's grant had previously been cut by 40 per cent when the present government took office last year, revealing a dramatic change in policy towards research.

Grants totalling FF930 million (compared to FF750 million for 1987) have been earmarked for the Fonds de la Recherche et de la Technologie (FRT), a research fund reserved for national priority research programmes. Valade

European data net under threat

London

PROPOSALS by the West German telecommunications authority, the Bundespost, to impose usage-related charges on top of the costs of leased data lines would, if implemented, have "disastrous consequences" for European academic research, Dennis Jennings, president of the European Academic Research Network (EARN), warned last week. Such action by an individual national authority would undermine strategic research programmes and would be against the best interests of the European community, Jennings says. Last week EARN decided to give up four of the five international leased lines that it runs into West Germany. □

announced the 11 rubrics for which tenders may be submitted (by private as well as public sector research groups): (1) biotechnologies, (2) nutrition, (3) medical research (including AIDS and retro-virus research), (4) human sciences, (5) technology and production, (6) electronics and information technology, (7) transport and civil engineering, (8) natural resources and the environment, (9) new materials, (10) new chemistry and (11) development.

Valade singled out research on high temperature superconductors as a prime example of an area where industry and the *grands organismes* should collaborate. Under the new materials heading of the FRT list of priorities, FF20 million (\$3.28 million), will be available, in addition to existing grants to public sector physics and chemistry laboratories. Peter Coles

Franco-German agreement to tap hot rocks

Paris

FRENCH and West German geological establishments have signed an agreement to collaborate on an exploratory geothermal energy project at Soultz-sous-Forêts, in the Rhine valley near Mulhouse in Alsace. The first phase of the project, which should be completed in 1989, will be undertaken by the French Bureau de Recherches Géologiques et Minières (BRGM) and the German Bundesanstalt für Geowissenschaften und Rohstoffe, at a cost of FF50 million (£5 million). The European Economic Community is to contribute FF14 million to the costs from its general research fund, the rest coming from national research budgets — FF16 million from France and FF20 million from West Germany.

The geothermal gradient, around 3 °C per 100 m, usually makes it impracticable to exploit the Earth's heat energy at the surface, but at Soultz-sous-Forêts, geological anomalies have resulted in temper-

atures of 150 °C in the underlying granite at a depth of only 1,800–2,000 m. At this temperature the energy can, in principle, be used not only to heat water pumped into the drill hole, but also to generate electricity, using heat-exchanges at the surface. As the Rhine itself forms the frontier between France and Germany in this region, both countries would be able to benefit from the energy if it can be harnessed, hence the decision to collaborate.

Similar projects at Los Alamos in the United States and in Cornwall in Britain encountered serious difficulties in recovering hot water at the surface, but depths are greater than those in Alsace (3,500 and 2,500 m, respectively) and because this part of Alsace was also the site of exploratory drilling for oil, its geology is well documented. The disused wells and derricks will be exploited for the initial seismic, hydraulic and corrosion studies comprising phase I of the geothermal projects. Peter Coles

Fifteen per cent solution to keep Britain in synchrotron club

London

BRITAIN'S reputation as a fickle collaborator on international science projects has been further reinforced after last week's meeting of the European Synchrotron Radiation Facility (ESRF) council in Grenoble, France. Nine of the ten member states were able to give provisional commitment on funding the facility. Embarrassed UK delegates could not say how much money Britain will be able to afford, or indeed if it will remain a partner. Britain has until the end of the year to come up with a subscription fee acceptable to the other members.

Originally, five nations were involved in ESRF's £3 million foundation study, each member contributing a sum roughly in proportion to its gross national product: France (38 per cent), West Germany (28), Britain (15), Italy (15), Spain (4). Since the foundation phase started in April last year, five more countries have joined the venture, Switzerland and a 'Nordic consortium' of Norway, Denmark, Sweden and Finland. The Nordic countries will together contribute 4 per cent, as will Switzerland.

At last week's meeting, the ESRF council formally approved the foundation phase report, which covers an 11-year period, 6½ years for the machine's construction and 4½ for experimentation, at a cost of £350 million. The ESRF council wants work to start next year.

The council will meet in November and again in December, when final levels of funding will be agreed. Britain's contribution according to a formula based on gross national product should be of the order of 18 per cent. The other partners consider 15 per cent acceptable, but might not be prepared to tolerate a disproportionately small subscription fee from what is still seen as a wealthy nation. Britain's position depends on the outcome of the government's public spending review in the autumn.

The Science and Engineering Research Council (SERC) is keen to see Britain in ESRF. The present difficulties have arisen principally because SERC failed to entice France to enter a collaborative arrangement on the spallation neutron source, ISIS, at the Rutherford Appleton Laboratory. SERC had hoped that France would agree to take a 20 per cent stake in the facility. Discussions are still proceeding along similar lines with Italy.

Synchrotron users in Britain are lobbying hard for inclusion in ESRF at the level being requested. As with most international collaborative ventures, contracts for building the machine will be awarded to individual countries in proportion to

their contribution, and the input and influence of the scientists using the facility will be similarly determined. Such advantages would be lost if Britain decided to join the facility once it had been built. Furthermore, ESRF's supporters say, the sums involved are, by comparison with other international projects, relatively small. At 15 per cent, Britain's contribution would be on average between £5 million and £6 million annually for the 11-year period, with the 1988 contribution possibly as low as £1.5 million.

As with Britain's involvement with the European Space Agency (ESA) and the European Organization for Nuclear Research (CERN), discussions on funding are plagued by uncertainty and patience on the Continent is wearing thin.

Meanwhile, a £1 million upgrading of SERC's own synchrotron source at Daresbury in Cheshire has been completed. The high brightness lattice is a modification of the 2 GeV synchrotron radiation source, increasing its brilliance by an order of magnitude. Inaugurating the new high brightness lattice on September 17, higher education minister Robert Jackson extolled the benefits of international collaboration on scientific projects. The irony of his remarks will not be lost on the beleaguered British officials at CERN, ESA and now ESRF. **Simon Hadlington**

Europe's half-hearted research budget

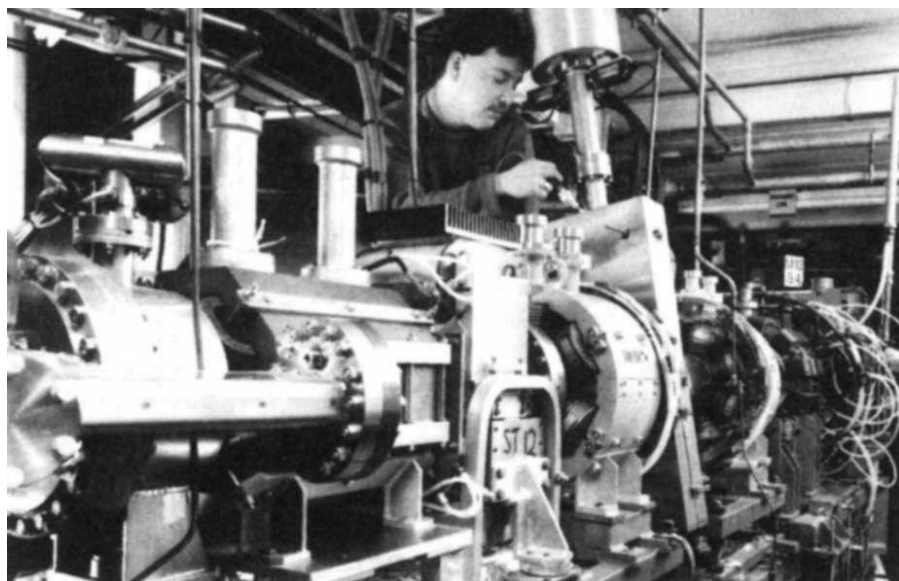
London

THE twelve Common Market governments were expected this week finally to approve the five-year Framework programme of research and development. The budget will be far leaner than the £7,300 million originally proposed, mainly at Britain's insistence that the programme in its original form was too costly and unfocused. The final figure is expected to be around £3,700 million, with the possibility of a further £300 million at the end of the year if community spending in other areas, notably agriculture, is brought under control to the satisfaction of all members.

Among the first casualties of the reduced research budget are more than half the projects selected as eligible for support under the latest phase of the BRITE (Basic Research in Industrial Technologies for Europe) programme. Last week 46 projects were selected for grant-aid worth £31 million. A total of 471 projects had been submitted for consideration, of which 250 were deemed eligible. A further 66 projects should receive £45 million after ratification of the Framework budget. More projects could have been supported had the original levels of funding been agreed, say officials.

The BRITE programme, launched in 1985 as a four-year project to increase the use of advanced technologies in traditional industry, now supports some 210 projects.

Simon Hadlington



THE construction and, more controversially, the financing of Europe's synchrotron sources has long been an area rife with international horse-trading. At an early phase in the negotiations to build the ESRF at Grenoble, the British partners (SERC) were keen to gain collaborators to share in the costs of the Spallation Neutron Source at Daresbury, near Oxford. French budget cuts last year effectively spelt the end to such hopes (Nature 321, 7; 1987), although now that the Daresbury source is in operation, the facility is open to experimenters from other countries who pay for time on the machine and related services. One of the refinements to the Daresbury synchrotron now upgraded is a smaller more concentrated beam, achieved by adding new focusing magnets (centre of photo above) to every straight section of the storage ring.

Scientists take share of blame for this year's poor Soviet harvest

London

THE lateness of the harvest this year is already causing heart-searching and rethinking among Soviet agricultural planners. One of the last major events of the Brezhnev era was the Food Programme, issued in June 1982, which was meant to ensure the long-term ability of the Soviet Union to feed its population by the year 2000. Although some lip-service is still paid to this programme, there now seems to be a tacit assumption that it will not materialize in its original form — and, indeed, one of its major features, the southward diversion of the north-flowing Siberian rivers, has been shelved.

Nevertheless, the main premise of the Food Programme persists — scientists should be more efficiently mobilized in the service of agriculture. A resolution of the All-Union Council of Ministers and Central Committee of the Communist Party of the Soviet Union in August took the same line. As part of the general “restructuring” of science on to a cost-efficiency basis, it offered financial incentives for research groups and institutes which contribute to increased productivity in agriculture. The profits from such innovations are, according to the new resolution, to be shared between the researchers and the farms which put the new methods into operation — thereby countering one of the main objections of scientists in past years, that workers on collective and state farms, who for some years now have enjoyed guaranteed wages, were not greatly interested in the changes and extra efforts involved in trying out new crops or procedures.

The August 1987 resolution, “on improving the scientific back-up for the country's agro-industrial complex”, has a more sophisticated approach to science than previous decrees. Instead of adopting the Brezhnev approach to the “chemicization of agriculture”, which advocated the use of pesticides and mineral fertilizers, it speaks of an “integrated” approach to pest control, using “ecologically safe” plant protection agents, including natural predators. Fertilizers are to be applied in accordance with growth and development, and there is to be more theoretical research into the mineral nutrition of plants. Instead of grandiose irrigation schemes the emphasis is on efficient use of existing resources, including drip-feed and synchronous pulse irrigation. Now more productive strains of livestock and plants are to be developed, including adaptations for the arid zones.

All this implies, that the scientists have been failing in their task — and *Pravda* said as much in a leading article on the new

resolution. The resolution “defined” the All-Union Lenin Academy of Agricultural Sciences as the “supreme” centre for the “comprehensive scientific back-up programme” but, *Pravda* said, it has failed to assume a leading role in the necessary experimental and implementation work. The All-Union Research Institute of the Electrification of the Agricultural Economy and the All-Union Research Institute of the Physiology, Biochemistry and Nutrition of Agricultural Livestock were singled out as examples of “low resultativity”. In particular, said *Pravda*, the new programme would necessitate increased efforts from the higher education sector, including major changes in training and updating courses and the establishment of special faculties for grading the qualifications of agricultural scientists and lecturers.

Two weeks ago a feature article in *Pravda* by Lenin-Prize winner G. Georgiev, which was outspoken even for the current era of *glasnost*, pinpoints the major legacy of Lysenkoism as an intense shortage of trained research personnel, a lack of proper appraisal of research results, and the absence of a small network of top-level laboratories (Georgiev spoke of 20–30) which would act as pace-setters for the rest. (What, one is left wondering, became of the massive emergency investment in molecular biology and genetics authorized in May 1974, which was intended to supply just such a base?)

Other scientists maintain that they have been working for agriculture on the lines now proposed for years, but that their results have been ignored or misapplied. In a recent Soviet television documentary, the president of the Moldavian Academy of Sciences, Dr Aleksandr Zhuchenko, spoke of major top-level errors in land-use, caused by politicians overriding the scientists' recommendations. In one case, he said, the All-Union Ministry of Land Reclamation and Water Resources had authorized plans for an irrigation system in the face of protests from scientists who pointed out that the water in question was unsuitable and would damage Moldavia's black soils. Zhuchenko further questioned the official Soviet approach to farm production targets. Instead of aiming at high yields in favourable years — and then writing off subsequent low results as due to adverse weather — the agricultural planners should aim at a balanced programme, producing good yields every year, including dry years, and should eschew doubtful management practices that can produce a one-off record crop, but from which the soil will then take ten years to recover.

Vera Rich

Fish psychology is big business

Tokyo

JAPANESE fisheries scientists are resorting to psychology to keep fish under their control. An experimental station in the southern island of Kyushu is applying electric-shock treatment to keep fish within a netless farm, while another is using the Pavlov technique and sound to confine red sea bream in an open bay.

Fish farming is big business in Japan with an annual production of just over one million tons — enough to satisfy the fish consumption of France. The electric fish farm is only in the very early stages of development. Hitachi Zosen Corporation, a shipbuilding company, and Kumamoto Prefectural fisheries station are testing the system in a tank. The farm is defined by three rings of metal stakes; the central ring is electrified. The distance between the central and outer ring is short, creating a strong electric field. Fish confined within the inner ring encounter an electric field that increases in strength outwards if they stray beyond that ring. Finally they are stunned before they can escape. When the fish regain consciousness on the sea floor they usually swim back to the inner circle. The electric farm avoids the problem of clogged nets which causes reduced oxygen supply and disease.

The sonic fish farm at Oita Prefectural fisheries station in Kyushu is at a much more advanced stage and employs quite the opposite psychological approach. Red sea bream are raised from eggs in tanks and then netted enclosures in the sea to the 3-cm stage. They are then conditioned to associate sound with food. Sounds are played to them at the time of feeding. At first a piano piece composed by an Oita University professor was used, but the fish did not like the variation in frequency. Bongo drums were not to their liking either. In the end, a plain ‘pooh pooh’ at 300 hertz and 50 decibels did the trick.

After 80 days of sound conditioning, the fish, now 10 cm long, are released into the bay where all the latest high-tech equipment is brought to bear to prevent the sea bream from swimming away.

One of the great advantages of the ranch is that the costs of feed and manpower are greatly reduced — 70 per cent of production costs of conventional fish farms is taken up by food and labour. And even although the 20 per cent recovery for released fish is less than that of a conventional farm (80–90 per cent), Yasumura estimates that production costs per kg for a sonic fish farm will be less. And as the fish look and taste like their natural wild cousins, he expects them to command twice the price of the conventional farm variety.

David Swinbanks

Acid rain report assailed as misleading

Washington

A NEW US government report on the causes and effects of acid rain is, depending on who you speak to, either a careful assessment of current scientific uncertainty or a misrepresentation of a serious problem. The report is the product of almost seven years of deliberation by scientists working on behalf of the National Acid Rain Precipitation Assessment Program (NAPAP), representing a number of government agencies. Its four volumes are acknowledged even by detractors to contain a comprehensive and mostly sound study of the sources of pollution, atmospheric chemistry, damage to lakes and forests, and potential methods of controlling emissions. But critics, such as Michael Oppenheimer of the Environmental Defense Fund, call the executive summary "skewed" in its summing up of the evidence.

Two particular disagreements stand out. Surveying all of the United States, the report declares that the number of lakes acidified to a damaging degree is no more than a few per cent of the total. NAPAP deems a lake unacceptably acidified only if its pH is 5.0 or less, when fish die and direct economic loss follows, but according to Jim Gibson of Colorado State University, there is obvious disruption to the food web, with concomitant loss of fish, at a pH of 6.0. The NAPAP report also emphasizes that no harm to seedlings has been demonstrated from acidic precipitation at "regional ambient levels", a statement which, says Oppenheimer, is true but misleading; extensive cloud cover of pH as low as 3.2 has been found in New England and it is these episodes, rather than average immersion at higher pH, that are responsible for forest damage in the north-eastern United States.

Many of those unhappy with the report see too much in it of the beliefs of NAPAP's chairman, J. Laurence Kulp, a geochemist. But Kulp defends the report, saying that the executive summary represents a consensus of the authors of each chapter and of senior scientists from all the participating agencies. He says that lakes naturally exist with a wide range of acidity, and that a pH of 5.0 is when biological activity falls off most sharply. David Streets, of Argonne National Laboratory, was "entirely comfortable" with the parts of the report to which he contributed. Bob Downing, a spokesman for Kulp's office, did admit that NAPAP had "learned the hard way" the perils of summarizing a hefty report in one-page chapter synopses.

Described by his opponents as "strong-willed", Kulp was apparently given his

position at NAPAP for strength of character rather than political leanings. But his panel's report carries undoubted political weight in the dispute between Canada and the United States over the effects of acid rain. Ross Glasgow, First Secretary for the Environment at the Canadian Embassy in Washington, described the NAPAP report as a "very unhelpful" contribution to the acid rain debate, and contrasted it with laudable US action on atmospheric ozone, where the scientific evidence is less clear. But so far there has been no government comment on the NAPAP report, and Glasgow believes that strong criticism of it within the United

States may already have undermined its credibility.

Canadian scientists were given a preview of the report before publication, and made known their worries that by neglecting the higher acidities in Canadian surface waters the magnitude of the problem was not fully expressed. And although, says Glasgow, the NAPAP report was specifically a US study, its omission of evidence from Canada and Europe lent bias to its conclusions.

Kulp himself is about to resign his chairmanship of NAPAP having completed what he regards as a "major milestone". NAPAP, established in 1980 with a ten-year mandate, will continue to compile evidence. A new chairman has not been announced, but both Kulp and Downing expect the programme to carry on in the same vein. **David Lindley**

Britain's universities outshone by US fund-raisers?

London

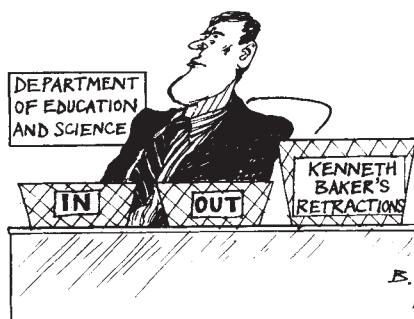
EMBARRASSED officials at the Department of Education and Science (DES) are trying to play down an off-the-cuff remark by education secretary Kenneth Baker which implied that British universities were not putting in enough effort to raise funds from private sources. Speaking in Chicago during a tour of the United States last week, Baker was reported on BBC radio as saying that more British universities should follow the example set by their US counterparts by employing full-time fund-raising staff. According to the report, Baker said that Southampton was the only British university to do this.

His remarks prompted an immediate and angry response from the universities. The Committee of Vice-chancellors and

fund-raisers. According to an official, Baker was merely airing his view that fund-raising was "a good thing that ought to be encouraged". The universities are aggrieved because they feel they have responded positively to the harsh economic climate by actively seeking non-government support. In 1985-86, Salford University earned nearly a quarter of its total income of £28.9 million from the open market. A spokesman from Kent University, which is currently seeking a Development Officer to launch a major fund-raising campaign in 1990, says that Baker's remarks will "reinforce public prejudice that the universities are lying on their backs waiting to be spoon-fed with government money".

Warwick University, which earned 15 per cent of its £37.3 million income from non-government sources in 1985-86, points out that it built the largest arts centre outside London with mainly private funds. London University's Imperial College has recently appointed a Director of Public Affairs, with responsibility for fund-raising. In general, most British universities do not have a strong tradition of raising funds from alumni, although it is a potential source of funds receiving considerable attention. University fund-raisers say that despite Baker's exhortations, funds from private sources will not materialize overnight. Furthermore, it is argued, the government itself could help by introducing more favourable tax incentives for corporate and charitable donations. Perhaps the biggest disappointment over Baker's alleged comments is that the Secretary of State is apparently unaware of the universities' efforts to do as the government has told them.

Simon Hadlington



Principals was quick to point out that nearly half of Britain's universities employ fund-raisers, and that universities have doubled their income from outside sources in four years. DES officials — also faced with objections from British school teachers upset at Mr Baker's apparent preference for US teaching methods — claim that Baker was inaccurately quoted and that he mentioned not only Southampton but also Oxford as two examples of British universities that have full-time

Chaos and complexity reign in the New Mexico mountains

Santa Fe, New Mexico

BARELY three years old, the Santa Fe Institute (SFI) can hardly be accused of having modest goals. "We are trying to define the scientific and educational agenda for the twenty-first century", says David Pines, University of Illinois physicist and co-chair of the SFI's science panel. The immediate challenge for the institute is to find funds among traditional sources for some non-traditional ideas.

From the outside, SFI is unassuming. It is located in a converted convent, a brief walk from the state capital, down a road famous for its art galleries and craft stores. But inside there have been inter-

visions of how the institute might evolve. Physicist Murray Gell-Mann of the California Institute of Technology is one of SFI's founders, serving now as co-chairman of the science board and a member of the board of trustees. He sees SFI as providing an alternative to universities where it is difficult to foster interdisciplinary interactions. Gell-Mann says interdisciplinary centres are often relegated to the basement of a university-owned but unused Victorian mansion, with staffs forced to survive on short-term grants, shunned by the academic departments to which the centres nominally belong.

SFI currently maintains a small permanent staff. If money is found to support expansion, the core will stay small, but there will be a coterie of visiting faculty and graduate and postgraduate students.

SFI has been surviving on foundation grants and private contributions. But word is expected shortly on two multi-million dollar grant applications, one to the National Science Foundation and one to the Department of Energy. There is apprehension among SFI officials that the grants may be difficult to get, since the institute does not fall into a convenient category.

The proximity of the Energy Department's Los Alamos National Laboratory has a significant influence on SFI. Many of SFI's board members are affiliated with the national laboratory, and SFI co-hosted with the Los Alamos Center for Nonlinear Studies a workshop on artificial life.

There is no unanimity among board members about the specific paths SFI should take. But those actively involved with the institute radiate an excitement about its potential, and they believe that through the interaction of talented people there may well arise new visions of how the world is put together. **Joseph Palca**



Gell-Mann — who says SFI should provide a nurturing environment for those looking to cross traditional boundaries.

esting goings on. Throughout the past summer, SFI has hosted a series of meetings and workshops to begin the process central to its goal of helping to catalyse new, interdisciplinary approaches to understanding complex systems.

A recent example was a meeting dubbed "Evolutionary paths of the global economy", co-chaired by physicist Philip Anderson of Princeton University and economist Kenneth Arrow of Stanford University — both Nobel prize winners. Sponsored by Citicorp and the Russel Sage Foundation, the meeting brought together economists and physical and biological scientists who have used nonlinear dynamic models to study adaptive paths.

Other intriguing topics discussed this summer include the matrix of biological knowledge — an attempt to bring the tools of computer science and artificial intelligence to bear on the increasing volume of biological and clinical data — theoretical immunology, and a meeting this week on computational approaches to evolutionary biology.

SFI president George Cowan, who is also a senior fellow at nearby Los Alamos National Laboratory, says there are many

Eastern bloc exchanges may benefit from new rules

London

THE inflexible regulations and complicated red tape that have made life difficult for Soviet scientists working abroad should be eased, the authorities claim, with the recent changes in Soviet rules for travel. Existing Soviet procedures for foreign assignments for scientists and technologists have come under strong criticism of late — both from a senior Soviet diplomat and the scientists themselves. Interviewed by *Izvestiya*, Strizhkov, the Deputy Economic Counsellor at the Soviet Embassy in Sofia, said that although 376 Soviet specialists are now established at Kozlodui, where they are officially assisting the Bulgarians in the construction of a nuclear power station, many of them are not required at the present time.

The Soviet practice is to send such persons abroad on a two-to-three years' posting; however, many of them are actually required on the site only for a series of relatively short periods, and, if the paper-work permitted, could easily return to the Soviet Union in between. Unfortunately, no one will send them home, since there is no guarantee that they will be able to return when required. And, if a particular specialist has completed his term at the site, he has to go back to the Soviet Union, even if the task for which he has come has not been completed.

Plans for the supply of Soviet specialists to such projects are laid down in advance and until now there has been no provision made for change of schedule or emergency. *Izvestiya* quoted a case where

problems arose with equipment supplied to Kozlodui by a Hungarian plant, a West German firm and the Soviet Tulaelektroprivod trust. Hungarian and West German trouble-shooters were on site within a week; the Soviet team arrived six months later. The latest changes, it is hoped, will make such "emergency visits" easier.

As well as keeping specialists idling abroad, long-term postings mean the need to support a considerable infrastructure: the Soviet colony at Kozlodui usually contains 800 to 1,000 persons, with inevitable problems of housing, transport, medical services, children's schooling and so on.

There is considerable pressure from the Kozlodui Soviet colony, *Izvestiya* said, for such "fraternal assistance" to be organized on a task-basis rather than a time-basis. Such a view-point is likely to receive considerable support from Soviet industrial managers. Until now, managers have been reluctant to have their specialists and scientific staff coopted, since although the host country paid for all technical services provided by the Soviet Union, none of this money filtered back to the enterprise which provided the specialists. Under the new Law on the State Enterprise, adopted on 1 July, all foreign "economic activity" including the "export" of consultancy and technical services, will be considered as part of its overall performance. Enterprise managers will therefore wish all such foreign "activity" to be planned on a basis of cost-effectiveness, and, *Izvestiya* argues, will favour any scheme which gets the specialist back to his or her regular job as soon as possible. **Vera Rich**

Hungary as neurosciences audit

SIR—The meeting of the International Brain Research Organization (IBRO) — the second World Congress of Neuroscience, held in Budapest from 16 to 21 August — with some 3,000 participants, makes an interesting testing ground for international comparisons on relative scientific standing. I have therefore analysed the oral and poster presentations listed in the conference programme by geography of origin. In the recent past British contributions to neuroscience have been pre-eminent, and one might have expected a high British profile at the meeting. Because the meeting was held in Hungary, it could be presumed that access was relatively easier to Europeans, East and West, and somewhat less to participants from other continents, than would have been the case with another venue.

The oral presentations were all invited, and therefore indicate the 'visibility' both of subject areas and the leading researchers within them. Of 612 listed papers, only 59 — just under 10 per cent — were from UK laboratories. Thirty-seven per cent were from East or West Europe, 37 per cent from the United States, a miserable 3 per cent from the Soviet Union and 13 per cent from the rest of the world, predominantly Japan, Canada, Australia and Israel.

Poster presentations were self-contributed, and as far as I know all those submitted were accepted. In one full day's set of 490 posters, just under 20 per cent of the entire meeting, only 23, or fewer than 5 per cent, were from UK laboratories. Thirty-one per cent were from East and 34 per cent from West Europe, 13 per cent were from the United States, fewer than 5 per cent from the Soviet Union, and the rest of the world contributed 12 per cent.

The first conclusion to be drawn is that the high visibility of contributions from US laboratories, leading to their domination of the oral sessions, can only enhance the well-known Matthew effect by which to him that hath more shall be given.

Second, the continuing prestige of senior UK scientists ensured them a significant presence among the oral presentations, but the weakness of the underlying position of UK neuroscience is shown by the small number of poster contributions. The more junior British scientists either had little to say or, more likely, were simply unable to raise the funds to attend. These two observations support the view that if European science is to increase its own visibility, it must reinforce its own cooperative mechanisms rather than stand so passively in awe of US achievements.

Third, the tiny representation of Soviet scientists even at an Eastern European meeting of a Unesco-sponsored organiza-

tion to which the Soviet Union is affiliated, and despite the distinguished Soviet neuroscientific tradition, may say something about how far *glasnost* still has to go.

Fourth, it is a mockery of the concept of internationalism when a 'world' meeting boasts no contributions from Africa and only a handful from India, China and Latin America. IBRO's world remains overwhelmingly white.

Fifth, the gendered pronoun I used to define the Matthew effect was no slip of the pen. I estimate that 20–30 per cent of those attending the meeting were women. A significant number of the posters had women authors and women presenters. But women speakers in the oral sessions were conspicuous by their absence. Given the chance to choose speakers for these sessions, it seems the predominantly masculine group of organizers and chairs could see only men.

The pattern of this meeting is likely to be mirrored in the thousands of other international meetings held each year. And it is precisely because of this — and also because in this case the organizing body, through its ultimate links with the United Nations, has legitimate claims to universalism — that the experience is worth pondering, rather than being regarded simply as part of the natural order of the scientific world. At the least, positive efforts, nationally and internationally, are needed to help redress the imbalances that this simple numerical exercise reveals.

STEVEN ROSE

*Brain Research Group,
Open University,
Milton Keynes MK7 6AA, UK*

Aleksandrov's visit

SIR—Vera Rich (*Nature* 327, 544; 1987) writes that SCOPE may have put restrictions on Vladimir Aleksandrov during his US visit in January 1985. Not so. SCOPE-ENUWAR has given every opportunity for speedy communication of new information, incurring criticism for not going to recognized journals with standard refereeing. I know Aleksandrov was not allowed direct access to Cray computers during his US visit.

Aleksandrov visited Japan in February 1985 for the SCOPE-ENUWAR workshop and said he was able freely to discuss climate modelling during his US visit. In Japan he openly met scientists from Livermore, Colorado State University, NCAR Boulder, NASA-Ames, France, West Germany, Australia, Sweden and Japan. He talked to me about personal problems of his wife's health and up to the time of his disappearance was sending medical reports for study in England with a view to arranging treatment.

Aleksandrov is still missed by his collaborators in ENUWAR, and at their Bangkok meeting in February they renewed the consensus on the predictions of climate change made in the SCOPE report which acknowledged the insight given by the limited one-dimensional model of Aleksandrov and Stenichkov.

FREDERICK WARNER
(Chairman, SCOPE-ENUWAR Steering Committee)

*Department of Chemistry,
University of Essex,
Colchester CO4 3SQ, UK*

VERA RICH REPLIES — The suggestion that Aleksandrov may have been restricted during his final visit to the Livermore Laboratory was originally made from the Soviet side, and confirmed in telephone conversations by his US hosts, who clearly felt awkward about the situation. I made no suggestions about the propriety or otherwise of such surveillance — which, according to Aleksandrov's hosts, was considerably less than that which I would accept as normal in working in a socialist-bloc country. □

Argentine trials

SIR—K.S. Jayaraman (*Nature* 328, 287; 1987), refers to a statement by Dr S. Ramachandran that "Argentine controversy (over rabies vaccine) arose because Wistar failed to inform the government".

The published facts (*Nature* 326, 636; 1987) reveal that: (1) The trial in Argentina was carried out by the Pan American Health Organization not by the Wistar Institute. And (2) the "controversy" arose because an Argentinian scientist decided to play politics instead of considering the trial as an important step in evaluating a vaccine which may represent the best way of preventing rabies infection.

HILARY KOPROWSKI

*Wistar Institute,
36th Street at Spruce, Philadelphia,
Pennsylvania 19104-4268, USA*

Human embryo

SIR—Your suggestion (*Nature* 327, 87; 1987) to ban the unnecessary word "pre-embryo" (an embryo before implantation) is well taken. At the Carnegie Collection, in human embryology, the term embryo is used for the "human offspring in first eight weeks" (*Concise Oxford Dictionary*). Hence prenatal life is subdivided into merely two periods: embryonic and fetal. Terms we have discarded for the human embryo include ovum (used incorrectly for everything from unfertilized oocyte to three-week embryo), 'egg' (best reserved for a nutritive object sometimes seen on the breakfast table), and blastula, blastocoel, blastopore, gastrula, branchial, all of which are inappropriate.

R. O'RAHILLY
*Carnegie Laboratories of Embryology,
California Primate Research Center,
Davis, California 95616, USA*

Physics successes

SIR—John Maddox made an unfortunate remark in his article (*Nature* 328, 375; 1987) on the meeting at Bristol to commemorate the 1947 discoveries of the pion and the V-particle. I did not detect in the nostalgia of the Bristol meeting any indication "that these developments should have turned out to be the last important contributions of British experimentalists to the field...". On the contrary, I believe they were simply the first.

British physicists have taken a prominent part in all the major experiments at the European Organisation for Nuclear Research (CERN) and Deutsches Elektronen-Synchrotron (DESY) over the past decades (University College London and Oxford in the discovery of neutral currents in 1973; Birmingham, the Rutherford Appleton Laboratory and Queen Mary College in the discovery of the W and Z particles in 1983; Manchester, Lancaster, Oxford and Imperial College London in the discovery of gluons in 1979). They also played a leading role in neutrino and muon scattering experiments at CERN over the past decade, identifying partons with quarks and providing the first quantitative support for the theory of interquark forces (quantum chromodynamics). Even the Kendrew committee paid tribute to the quality of their work.

Because the people concerned work in large international teams, the role of individuals is not as clear as it was in 1947. But to dismiss the post-1947 achievements of the UK particle physics community is indefensible and just plain wrong.

D.H. PERKINS

University of Oxford,
Department of Nuclear Physics,
Keble Road, Oxford OX1 3RH, UK

Geology departments

SIR—The University Grants Committee's Report of the Earth Sciences Review concludes that resources in this area in the United Kingdom are too thinly spread and that significant benefits would result from their concentration into a smaller number of large departments. Although several reasons are given why research productivity might be favoured by large size no firm evidence is presented. To shed light on this issue, I analysed the published research output and citation rates of 33 university departments of geology or earth sciences in UK universities.

An indicator of the quantity of published work of each department was derived from the *Science Citation Index* (SCI). Citations are collected from a full coverage of every item published in 3,250 journals as well as selective coverage of other publications. The Corporate Index volumes of SCI provide a convenient

source of information classified by university and department. A count of substantial items was made for the three years 1982 to 1984.

The information on the volume of publications was supplemented by an analysis of citations, allowing distinctions to be drawn between work that has exerted a significant impact on a research field and that which has not. As citations are listed by individual, a list was made of names of individual members of each department as given by the *Commonwealth Universities Yearbook* for 1985; only full-time academic staff were included. For these individuals, citations in SCI were counted for the period used in the analysis of publications. Information on number of faculty in each department was also used to calculate *per capita* rates for publications and citations and as the indicator of size.

The correlation coefficient for the relationship between size of department and numbers of publications per member of faculty was 0.112 and that with number of citations per member of faculty was 0.256. Neither is significant at the 0.05 level. The results therefore provide no evidence of any significant relationship between research productivity (as measured here) and departmental size. Whatever the other merits of a concentration of resources, the view that this would enhance research productivity should be treated with some scepticism.

GRAHAM BENTHAM

School of Environmental Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK

South African view

SIR—As a South African expatriate I am struck by the muddled thinking in the recent correspondence over an academic boycott. Analyses such as those of J.G. Wilson (*Nature* 328, 288; 1987) and Max Wallis (*Nature* 328, 374; 1987) lead nowhere except back to their original ideological premises.

Like many South African-born families lucky enough to have the choice, my family left (in 1975) when it seemed that, whatever political options were chosen, there was no acceptable solution. While opposed to apartheid in principle, and abhorring many of its practical evils (the separation of black families, imprisonment without trial, the ultrahypocritical actions of many whites), we saw no political alternative to which to aspire.

The simplistic approach of 'one man, one vote' was not tenable. Like it or not, lumping together large groups of people with diverse cultures creates a sure-fire setting for political nightmares, as is evident in many of the world's worst trouble-spots. Although W.D. Stein (*Nature* 328, 374; 1987) claims this would not be as catastrophic for blacks as for whites, the

history of independence in Africa suggests otherwise. This is not to argue that majority rule should be implacably resisted — it is only wise, when faced with the inevitable, to prepare for it. One must, however, be an especially confident ideologue to demand particular changes regardless of the trauma they invite.

These feelings support two conclusions: (1) there are no easy options for the white electorate in South Africa and (2) boycotts will not force whites to choose political platforms that serve no one except the next tribally based clique to assume the reins of government.

The question for outsiders thus boils down to a personal view of how to ease the tribulations that South Africans of all colours must undergo in the transition to majority rule. At present, this means adopting tactics that temper, as far as possible, the more brutal and reactionary aspects of the Nationalist regime. It argues for a strong dose of pragmatism.

Boycotts have had a significant political impact on South Africa. In sport, they may have done some good by forcing white electors to recognize the repugnance with which apartheid is viewed by nations they would call friends. In trade and economic cooperation, the issue is more difficult because black workers are more immediately affected than whites.

But what can be achieved by an academic boycott? The state the universities in South Africa, as in most places, is of virtually no concern to people in the street. They could hardly care less about the scientific well-being of academics. Neither, therefore, do the politicians care a rap. Nor will a boycott by scientists prevent South Africans from obtaining the technology they require. Yet the damage to liberal thinking in South Africa by isolation of the universities would be profound.

PETER M. MAY

School of Mathematical and
Physical Sciences,
Murdoch University,
6150 Western Australia

SIR—While it is probably inadvisable for a resident to comment on the views expressed in the recent correspondence following your article on the academic boycott of South Africa, I should like to record that in my second-year organic chemistry class this year there are 24 Indians and 5 Africans, the balance being 52 of European origin.

A few years ago, the class was of a similar size, but completely European. Although we are not allowed to increase the size of our intake, the racial composition is changing rapidly.

D.A.H. TAYLOR

Department of Chemistry
and Applied Chemistry,
University of Natal,
Durban 4001, South Africa

The Antarctic cornucopia

New results from a Soviet–French collaboration highlight the benefits to be gained from polar research, as well as a US scientific deficit.

THOSE who have spent time in Antarctica do seem to be something of a breed apart, not least because they generally return to 'civilization' itching to depart southward again at the earliest possible opportunity. This phenomenon must stem from some personal experience denied to less hardy mortals — the memory of a brilliant landscape of white and blue beneath an incomparably clear atmosphere, perhaps, or the attainment of as perfect a silence as nature will allow. A more mischievous speculation would dwell on the breathing space for misogynists provided by those national surveys that have up to now been staffed exclusively by males (though even the British Antarctic Survey may remove that particular iniquity soon).

Such superficial attractions are easily matched by other considerations. The political advantages of Antarctic research were amply highlighted by the sudden conversion to the cause by the British Prime Minister following the Falklands war. Other attributes of the region are self-evidently attractive to scientists: the Antarctic circumpolar currents which serve as both source and sink to deep ocean flows, of potentially crucial significance for climatic change; the proximity to an uppermost atmosphere linked directly to interplanetary phenomena by a downward-dipping geomagnetic field; a unique ecosystem; and, of course, a stratosphere whose dynamics and chemistry have now shone a red light at the chlorofluorocarbon industry.

But as a source of fascination the ice sheets are hard to beat. Their ponderous dynamics are unusual enough — the coupling between gravity, thermal flow and temperature-dependent rheology has so far rendered it impossible to solve several fundamental problems. For example, might a combination of rapid accumulation and bottom melting lead to a catastrophic surge? Is there a natural resonance within the combination of ice sheets, oceans and atmosphere, that can account for the strong glacial response to the tiny variation of incoming solar radiation caused by the 100,000 year Milankovich orbital ellipticity cycle? Progress is hindered by the effort it requires simply to measure in the field such fundamental quantities as ice flow velocities.

Such experimental and theoretical problems posed by the ice sheets have not, of course, prevented their fruitful exploitation, as three Franco–Soviet papers on

the Vostok ice core in this issue amply demonstrate (pages 403–418). The flow properties are sufficiently well known, and the past annual snow accumulations sufficiently well inferred, for it to be possible to date ice (albeit with significant uncertainties) as one probes downwards. And as snow is compacted beneath successive falls, air within it eventually becomes trapped in bubbles. So if due allowance is made for the several decades such entrapment takes, one in principle has a direct chronological record of atmospheric composition and — thanks to stable isotopes and their various fractionations — of climate stretching back as far as the sheets permit.

The Vostok core is the deepest to have been drilled. Its 2,200 metres extend back over 160,000 years through the ice age that ended about 8,000 years ago and into the previous one. Of course, deep-sea sediment cores can extend over a million years, but they can be far harder to interpret given the combined influences of climate and oceans on the indicators that they contain. As it is, the new Vostok record of carbon dioxide and temperature provides crucial information for those wishing to assess the relative importance of greenhouse effects on past climates, as Eric Sundquist explains on page 389.

The engineering skills required for deep ice drilling are too easily forgotten. Conventionally, drills have used rotating blades around the cylindrical sampler base to cut into the ice. At Vostok the Soviets have perfected a thermal technique whereby the base of the 8 m × 10 cm tube is electrically heated, penetrating the ice as it melts without damaging the sample.

As one might imagine the technique is not without its difficulties, quite apart from those arising from the Antarctic weather. Simply keeping the hole vertical requires carefully controlled descent, while the meltwater produced by heating has to be rapidly recovered to prevent refreezing and distortion. Should the drill become jammed for any length of time the inexorable flow of ice will deform the hole, making retrieval impossible. And complications arise from the need to balance glaciostatic pressure by filling the hole with a fluid, such as alcohol, which can withstand the –55°C average surface temperature.

Small wonder, then, that several holes were attempted (as can be seen on the cover of this issue) and that the core took

from 1980 to 1985 to be extracted. The current hole has now had to be abandoned but a new hole is to be started soon, potentially reaching a further kilometre or more into the ice. At such depths an annual layer may be only 1 cm thick.

A look at the background of such drilling might repay a historian of Pleistocene and Holocene science. The range and the grandiose implications of the data that can be retrieved from ice and sea cores tend to eclipse the inevitably more gradual progress made by those concerned with continental records. This seems to have been mirrored by the comparative support for the respective communities, at least in France — a situation that seems unlikely to change given the selective policies adopted by the committees responsible for allocating money.

In the international context one can only express surprise at the low profile recently achieved in this area by the United States. Most glaciologists would identify Willi Dansgaard of Denmark, Chester Langway of the United States, and Hans Oeschger of Switzerland as the prime innovators of glacial palaeostudies, with US expertise in drilling being particularly crucial in the 1960s for the path-breaking Camp Century core in Greenland, for example. But although the United States has continued to support drilling activities of its own and of other countries, scientific returns within the US community have not maintained the early prominence, and Western Europeans, particularly from France, Denmark and Switzerland, now dominate the literature.

No doubt the National Science Foundation and the community share responsibility for this odd state of affairs, and the foundation is quick to point out that a major programme of Antarctic and Greenland drilling is about to commence. The danger is that such a programme may lack a sense of focus, given the diffuse character of current US involvement. Two factors may be particularly relevant. In each of the successful European countries, efforts are concentrated within multidisciplinary glaciological laboratories. And two particularly successful international enterprises — that involving the three individuals above on the one hand, and the French–Soviet work at Vostok on the other — originated in, and have succeeded as a result of, close but purely informal collaboration.

Philip Campbell

Archaeology

Return of the Euro-boomerang

Paul G. Bahn

PAWEŁ Valde-Nowak and colleagues on page 436 of this issue¹ report the discovery of a curved piece of mammoth tusk in a cave in southern Poland. They claim that the tusk is about 23,000 years old, and that it is the oldest boomerang in the world.

In the minds of children and adults alike, the boomerang is considered as uniquely Australian as the kangaroo or the vegemite sandwich. It has certainly been part of the Australian scene for a long time, as shown by the three complete wooden specimens from Wylie Swamp, South Australia, which are 9,000 or 10,000 years old².

But boomerangs and 'kylies' (killing-sticks) have been found on five continents: some cave paintings in North

what do you call a boomerang that doesn't come back? A stick — is of relevance to this discussion. Killing sticks are far more numerous than boomerangs in both prehistoric and historic times, and probably predate the boomerang. The chances are high that the remarkable aerodynamic properties of sticks which happened to be curved were noticed and exploited independently on many occasions.

The only way to be sure that a curved object is a boomerang is to try it out, but few scholars would contemplate hurling prehistoric specimens into space to see if they come back! The best method is therefore to experiment with a cast or a replica, as has been done successfully with a plywood copy of the Dutch specimen, and with a replica of one from the eleventh Dynasty of Egypt⁴. But there is always the

problem that many such prehistoric objects have been warped or damaged during the time they were buried.

No such experiment has yet been done for the Polish find, and meanwhile Valde-Nowak *et al.* are basing their claim on its shape, its curvature and the flattening at both ends. It has a span of 70 cm, is up to 6 cm wide, and up to 1.5 cm thick (see their Fig. 1 on page 437). One side preserves the external, convex tusk-surface, whereas the other has been polished almost flat. On the evidence of morphology, therefore, the object certainly makes a plausible boomerang, and the find is of potentially great importance for our knowledge of palaeolithic techniques of exploiting game, as killing sticks, for example, can be accurate up to about 200 m, much farther than a man can throw a stone or spear³. □

1. Valde-Nowak, P., Nadachowski, A. & Wolsan, M. *Nature* 329, 436–438 (1987).

2. Luebbers, R. A. *Nature* 253, 39 (1975).

3. Bell, J. *New Scientist*, 99, 839 (1983).

4. Hess, F. *Antiquity* 47, 303 (1973).

Paul G. Bahn is a freelance writer of archaeology at 428 Anlaby Road, Hull HU3 6QP, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Pitt Rivers Museum, Oxford

Ancient Egyptian throwing sticks or boomerangs (from a, XII Dynasty Thebes; b, XII Dynasty or later). Scale bars, 5 cm.

Africa, for example, dating to perhaps 7,000 BC, are thought to depict an object of this kind, and boomerangs with gold caps at either end were found in Tutankhamun's tomb, dating to the mid-second millennium BC. Some form of throwing stick is also known to have been used by groups as far-flung as Eskimos, Hopi, Polynesians and Indians. In most regions, the invention of the bow and arrow eventually made the technique obsolete³.

In Europe, there have been finds both from the Iron Age (an oak specimen of 300 BC, preserved in a bog in Holland⁴) and from the Mesolithic (a find in Jutland from a settlement of about 5,000 BC). A piece of mammoth bone from an Austrian site was claimed to be a fragment of an Upper Palaeolithic boomerang, but its date and its identification are far from certain.

Valde-Nowak and colleagues think their new find, from a cave in the Oblazowa Rock in southern Poland, is a complete boomerang. It was well stratified with animal remains and with stone and bone tools attributed to the Pavlovian culture.

Not all curved objects, however, are necessarily boomerangs. The old joke —

Palaeoecology

Chalk grasslands in the ice age

Peter D. Moore

FOR 50 years, studies on the history of the British landscape have shown that there are no habitats that have escaped the influence of human activity, either directly or by the grazing of domestic animals. Indeed, many of the most wild and remote habitats that might have been considered relatively insulated have turned out to be the direct product of human management, including moorland, heathland, blanket mires, many valley mires and montane grassland. Research on the history of the chalk and limestone grasslands of Britain has been hampered by a paucity of evidence, but it is generally assumed that these areas were wooded and that former woodland was cleared by Neolithic or subsequent cultures over the past 5,000 years or so. On page 434 of this issue¹, Bush and Flenley describe the results of pollen analysis of sediments in the Yorkshire wolds, where the chalk exposures of England reach their northernmost limit. The authors show that grassland in this region persisted through the early part of this Flandrian interglacial, 10,000–8,000 years ago.

Chalklands present several problems for the palaeoecologist. The porosity of the substrate means that lakes and peat deposits are scarce or absent, so there are no stratified sediments for fossil analysis. In Britain, the chalk lies in the lowlands of the south and east where precipitation is generally low, so waterlogging and peat

formation become even less likely. Spring lines close to the chalk may bear wet woodland, particularly of alder, but when fed by calcareous water and subject to periodic drying and aeration, the preservation of microfossils like pollen grains is generally poor. Similarly, soils contain little pollen, but the lime-rich shells of gastropods do survive and provide valuable information about immediately local habitats. Much of what is known about the history of the English chalk is based on the evidence from fossil snails, both in late-glacial times² and during the periods of prehistoric forest clearance^{3,4}.

Pollen evidence concerning the history of vegetation on the chalk is not wholly lacking, however; information is currently available from the Isle of Wight off the south coast of England and from Sussex in the south of England. Scaife's work on the Isle of Wight⁵ shows that the open grasslands of the late-glacial were invaded by juniper and birch woodland soon after 10,000 years before present. The records in Sussex^{6,7} do not go back very far and are difficult to interpret because of the abundance of *Alnus* (alder) pollen in the stratigraphy. Because alder is likely to have been a local tree of the valleys, it confuses the picture of vegetation on the upper slopes of the downs.

Bush and Flenley have now discovered a site in Yorkshire where the sediments date back to the end of the last glaciation

and the early part of the current (Flandrian) interglacial. Unfortunately, there is a break in deposition after about 8,000 before present and sedimentation commenced again at about 4,300 before present, so the complete Flandrian history of the site remains unknown. But the record of the early part of the Flandrian clearly demonstrates the survival of open grassland right up to the discontinuity. Such persistence of grassland brings to mind the old theories of Wooldridge and Linton⁸, who considered chalk grassland to be a natural climax system that had remained free of forest throughout the Flandrian. But the evidence of Bush and Flenley cannot be taken to imply the absence of human influences. Indeed, the erratic behaviour of some trees, such as birch, and the occurrence of some weed species in the pollen rain, lead the authors to propose that Mesolithic people were already disturbing the habitat and maintaining open conditions.

As well as the long-term survival of grassland fragments in general, there is also evidence in these sediments for the specific survival of certain open-habitat species that are taking advantage of disturbances in the forest canopy. The discovery, for example, of *Teucrium botrys* is of particular significance, as this plant has been proposed as a survivor in chalkland refugia. Rose⁹ has accounted in this way for its presence on the chalk cliffs at Box Hill in Surrey — a classic refugium site where several relict species are found together. Direct palaeoecological evidence for the persistence of grassland refugia of this sort has previously been restricted to limestone sites in upland regions, such as Malham in Yorkshire, Upper Teesdale and Cwm Idwal in Snowdonia, Wales, but the work of Bush and Flenley adds weight to the hypothesis¹⁰ that many rare and disjunct species in Britain have become restricted in distribution by the spread of postglacial forests.

So, some fragments of Britain's chalk grasslands may date back to the close of the last glaciation and may have offered appropriate conditions for the survival of relict species, but this result does not imply that the grasslands can claim independence from the ubiquitous influence of humans. □

1. Bush, M.B. & Flenley, J.R. *Nature* **329**, 434–436 (1987).
2. Kerney, M.P., Brown, E.H. & Chandler, T.J. *Phil. Trans. R. Soc. Lond.* **B248**, 135–204 (1964).
3. Burleigh, R. & Kerney, M.P. *J. Archaeol. Sci.* **9**, 29–38 (1982).
4. Moore, P.D. *Nature* **299**, 780–781 (1982).
5. Scaife, R.G. thesis, Univ. London (1980).
6. Thorley, A.J. *Biogeogr.* **8**, 93–106 (1981).
7. Smyth, C.T. *Quat. Stud.* **2**, 22–33 (1986).
8. Wooldridge, S.W. & Linton, D.L. *Antiquity* **7**, 297–310 (1933).
9. Rose, F. in *Chalk Grassland* (eds Jermy, A.C. & Stott, P.A.) 6–9 (Trust for Nature Conservation, Kent, 1973).
10. Pigott, C.D. & Walters, S.M. *J. Ecol.* **42**, 95–116 (1954).

Peter D. Moore is Reader in Ecology at King's College London (KQC), Campden Hill Road, London W8 7AH, UK.

Palaeoclimatology

Ice core links CO₂ to climate

Eric T. Sundquist

FOR several years, researchers have struggled to explain the evidence of a 40–50 per cent increase in atmospheric CO₂ associated with the end of the last glacial period 10,000 years ago. This evidence, mostly from polar ice cores from Greenland and Antarctica¹, has sorely tested efforts to separate CO₂ and climate into cause and effect. One difficulty has been the short time-span of the ice-core data, which have extended back only 30,000 years and thus did not support conjectures about earlier Pleistocene warm periods. In three papers on pages 403, 408 and 414 of this issue^{2–4}, a Franco-Soviet collaboration presents ice-core results from the Vostok station in East Antarctica that show a remarkably close association between CO₂ and climate variations extending through the last interglacial period to the preceding glaciation about 160,000 years ago.

In one of these papers, Jouzel *et al.*² report the use of the deuterium/hydrogen ratio, measured in the ice at 1-metre intervals throughout the 2,083-metre core, to calculate a continuous record of changes in atmospheric temperatures. They estimate that the temperatures ranged from 9 °C below the present mean annual surface temperature of –55.5 °C (during the most recent and penultimate glacial maxima) up to 2 °C above (during the warmest part of the last interglacial). These conclusions are supported by an analysis of regional isotope-fractionation mechanisms, and by independent estimates from ice-crystal sizes, climate modelling, and temperature-related variations in ice accumulation rates. Even if these temperature estimates become somewhat modified by findings in the future (Jouzel *et al.* estimate the accuracy of the above departures to be about 20 per cent), they appear to be representative of a large portion of the Antarctic continent, and they stand as a powerful constraint for global climate models.

The CO₂ data reported by Barnola *et al.*³ in another of the papers in this issue not only confirm the deglacial CO₂ increase observed in previous ice-core studies, but also support the inclusion of CO₂ in general theories of Pleistocene climate change. The record shows two dramatic shifts, from about 190 to 280 parts per million by volume, that coincide with the ends of the most recent and the penultimate glacial periods. Between these features, the profile fluctuates between intermediate and low CO₂ concentration, corresponding generally to climatic conditions during the ice ages and interstadial

periods (brief warmings within an ice age). Other studies⁵ have confirmed, by overlap with the historical record of direct measurements of the atmosphere, that ice cores give a reliable indication of atmospheric CO₂ concentration. The Vostok CO₂ record appears to be irrefutable evidence for a fundamental link between the global climate system and the carbon cycle.

Like the recent less detailed oxygen-isotope record from the same core⁶, the Vostok deuterium and CO₂ profiles resemble the shape of the well-known oxygen-isotope curve from deep-sea sediments deposited throughout the world at the same time. (Compare Fig. 2 from Barnola *et al.* on page 410 with curve *c* of Fig. 2 from Jouzel *et al.* on page 404.) The latter is mainly a record of global ice volume. Its spectral characteristics have been shown to support the Milankovitch theory of climate change induced by cyclic variations in the Earth's orbital and rotational configuration with respect to the Sun⁷. Jouzel *et al.*² demonstrate that the Vostok deuterium record shows variance density peaks corresponding to periods near 40,000 and 20,000 years, values associated with oscillations in the tilt and precession of the Earth's rotational axis. In contrast, the spectral analysis by Barnola *et al.*³ of the CO₂ record does not yield a well-defined 40,000-year variance peak. Thus, although both the deuterium and CO₂ records suggest that astronomical forcing is involved, the role of this process is not simple.

Part of the difficulty in relating the Vostok data to other climate records stems from the long-standing problem of correlating ice and deep-sea stratigraphies. Much to their credit, the Vostok investigators do not assume a particular correlation in assigning ages to depths in their core. Instead, they rely on an independent age-depth curve derived from a two-dimensional glaciological accumulation and flow model. This scheme is sparsely documented, and has drawn fire from marine stratigraphers who argue that the Vostok chronology places the onset of the last interglacial period about 10,000 years too early, by comparison with dates assigned to the contemporaneous marine oxygen-isotope stage 5e.

Jouzel *et al.*² respond by citing ice-core measurements that support their dating, and point to ambiguities inherent in dating the deep-sea record by correlation with dated reef terraces. A recent analysis of diatom assemblages in a southern ocean sediment core (J. J. Pichon, personal communication) offers the prospect of

resolving the correlation problem by matching regional temperature curves. The debate is likely to intensify as a new oxygen-isotope profile, from a uranium-series-dated calcitic vein in the Great Basin in the United States⁸, places the beginning of the last interglacial stage even earlier than the Vostok date. These stratigraphic issues have profound implications for determining the timing and mechanisms of global climate change.

The Vostok data will also figure prominently in debate on which processes control atmospheric CO₂ concentration. The Vostok CO₂ data conflict with estimates based on a previous interpretation of marine carbon-isotope records⁹. Nor does the Vostok record confirm the very abrupt CO₂ fluctuations observed in the Greenland ice record¹⁰, although Barnola *et al.*³ emphasize that their sampling might not be adequate to reveal such changes. Deuterium and CO₂ appear to change almost simultaneously in the Vostok record of warming events, whereas in several cooling events, CO₂ changes lag significantly behind deuterium changes. Barnola *et al.* point out that different mechanisms seem to be required to account for these different modes of climate and CO₂ change. Although they focus on the influence of shallow versus deep modes of ocean circulation change, it is clear that mechanisms involving marine sediments and the terrestrial components of the global carbon cycle should be further considered.

Perhaps the most significant conclusion from the new results is that the global climate system and carbon cycle are intensely interactive. From a statistical

comparison of the Vostok temperature profile with linear combinations of the CO₂ profile, Northern and Southern Hemisphere insolation (solar radiation) curves, and several global ice volume profiles, Genthon *et al.*⁴, in the third paper in this issue, conclude that CO₂ is a dominant climate forcing factor. This result is not surprising, because it arises essentially from the close resemblance of the Vostok temperature and CO₂ curves. The authors emphasize that their approach does not exclude the reverse influence of climate on atmospheric CO₂.

Most important, Genthon *et al.*⁴ express the emerging consensus that CO₂ is an important component in the system of climatic feedbacks that amplify the relatively weak, direct effects of insolation changes. Similarly, carbon-cycle researchers continue to explore the various ways in which atmospheric CO₂ is sensitive to climate change. Further progress requires treating climate and the carbon cycle as parts of the same global system rather than as separate entities. □

1. Berner, W., Oeschger, H. & Stauffer, B. *Radiocarbon* 2, 227–235 (1980).
2. Jouzel, J. *et al.* *Nature* 329, 403–408 (1987).
3. Barnola, J.M. *et al.* *Nature* 329, 408–414 (1987).
4. Genthon, C. *et al.* *Nature* 329, 414–418 (1987).
5. Neftel, A. *et al.* *Nature* 315, 45–47 (1985).
6. Lorius, C. *et al.* *Nature* 316, 591–596 (1985).
7. Imbrie, J. *et al.* in *Milankovitch and Climate* Vol. 1 (eds Berger, A.L. *et al.*) 269–305 (Reidel, Dordrecht, 1984).
8. Winograd, I. *et al.* *Geol. Soc. Am. Abstr. Progr.* (in the press).
9. Shackleton, N.J. & Pisias, N.G. in *The Carbon Cycle and Atmospheric CO₂* (eds Sundquist, E.T. & Broecker, W.S.) 303–317 (Geophysical Monograph 32, Am. Geophys. Un., Washington DC, 1985).
10. Stauffer, B. *et al.* *Ann. Glaciol.* 5, 160–164 (1984).

Eric T. Sundquist is at the US Geological Survey, 431 National Center, Reston, Virginia 22092, USA.

Molecular dynamics

Simulating macroscopic flows

David Heyes

FLUID-FLOW patterns, caused by thermal and pressure gradients, are at the very heart of oceanography, meteorology and rheology (the study of liquids subjected to compressive and shear stresses). On page 427 of this issue, M. Mareschal and E. Kestemont report a numerical simulation of the convection currents caused by thermal gradients. Of particular interest is the authors' demonstration that these currents can occur on a microscopic scale.

Until recently the only way of predicting these flow fields was to postulate the guiding mechanisms of flow, for example, by using equations based on the continuity of mass and momentum. Apart from the numerical problems of solving often 'stiff' equations, these formulations are simply unrealistic at the high temperature gradients and stress fields that are encountered in, for example, shock studies

and non-newtonian liquid flows.

Essentially exact descriptions of flow patterns are nowadays achieved by generating them directly from the highly cooperative motions of individual molecules. This is possible using the computer simulation technique of molecular dynamics. The microscopic foundations of material properties are explored by generating the motions of the individual molecules from the forces that act between them. An assembly of model molecules is placed in a fixed volume, and numerical integration of the equations of motion of the molecules produces a series of fluid structures at time intervals from which the macroscopic properties can be derived. To avoid surface effects, the cell is surrounded by identical images of itself, to give periodic boundary conditions.

During the 1960s, molecular-dynamic

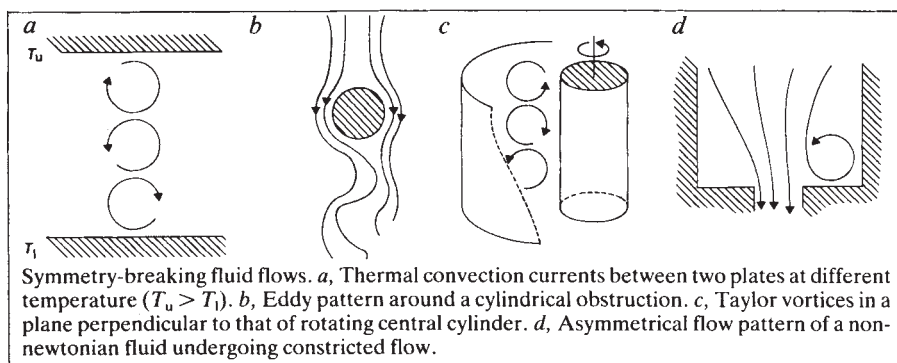
simulations on hard spheres ('snooker-ball' approximations to molecules) revealed that hydrodynamic flow patterns around a moving test particle could be reproduced by the Navier–Stokes equations, which therefore were shown to be valid on a microscopic distance scale (see, for example, Alder, B.J. & Alley, W.E. *Physics Today* 56–63; January 1984). Recent simulations of fluid flow past an obstruction (Rapaport, D.C. & Clementi, E. *Phys. Rev. Lett.* 57, 695–698; 1986) and convection currents induced by a temperature gradient (Mareschal and Kestemont in this issue) show that flow patterns such as eddies and thermal rolls (see figure) can be produced in calculations of cells just 30 nanometres across.

These and earlier simulations of more homogeneous fluid flow (see, for example, Ashurst, W.T. & Hoover, W.G. *Phys. Rev. A* 11, 658–678; 1975) reveal that physical properties of liquids remain unchanged when subjected to extreme perturbations. The viscosity, for example, has the same value up to enormous differential flow rates, of the order of thermal velocities across molecular dimensions. Similar microscopically large temperature gradients have an undetectable effect on the thermal conductivity. The macroflow simulations have made use of this 'fortuitous' behaviour to generate statistically significant results. The molecular-dynamics technique is also ideally suited to explore the onset of nonlinear material behaviour, simply by increasing the flow rate or temperature gradient.

Thus, calculations approaching the problem from both ends show that real macroscopic fluid behaviour can occur in systems sufficiently small to be handled by molecular dynamics. Such calculations are normally restricted by the number of individual particles the computer simulation can follow (160,000 in these studies). Both of the recent studies were restricted to two dimensions because of the shortage of computer power. Also they used low values of the Reynolds and Rayleigh numbers (25 and 1,780, respectively) that characterize the stability of a fluid against turbulent flow and convection. These values are comparable to those achieved experimentally because all dimensions, the size of the obstruction and the number of molecules are scaled down drastically in the simulation.

Despite these restrictions, both the studies generate time-dependent flow structures that would have been difficult to obtain with conventional numerical continuum models. If it is possible to use larger Reynolds and Rayleigh numbers, simulations of Taylor vortices (c in the figure) and Rayleigh–Bernard convection will be possible, as well as chaotic behaviour and the onset of turbulence.

Such calculations will not be easy, because the distance scale of the flow



pattern must be smaller than the smallest dimension of the model cell to avoid distortions caused by the periodic boundary conditions. Therefore, the size of the cell and the number of particles must increase with the Reynolds and Rayleigh numbers. The computer time increases at least in proportion to the number of particles considered, and the real time during which the particles are followed is typically less than 1 nanosecond.

Despite these temporary limitations,

molecular dynamics offers the ability, on one level, to generate molecularly based constitutive equations as a basis for macroscopic-flow calculations using finite-difference or cellular-automata algorithms (see Wolfram, S. *Nature* **311**, 419; 1984). In the long term, it could be used to investigate these phenomena directly. □

David Heyes is in the Department of Chemistry, Royal Holloway and Bedford New College, University of London, Egham TW20 0EX, UK.

Nonlinear dynamics

Chaos is good news for physics

Tomas Bohr and Predrag Cvitanović

In a recent paper, E.G. Gwinn and R.M. Westervelt (*Phys. Rev. Lett.* **59**, 157–160; 1987) explore a route from regular to chaotic behaviour in electronic transport in cooled p-type germanium. The same route to chaos has also been explored by A. Libchaber and co-workers in a very different physical system: the convective flow of mercury (Jensen, M.H. *et al. Phys. Rev. Lett.* **55**, 2798–2801; 1985). The new experiment is interesting as a study of semiconductor physics and for its many technological ramifications; but both experiments illustrate the high precision with which experimentalists can now test the theory of transitions to chaos.

Chaos has become a cover name for an active branch of physics, describing the highly irregular, unpredictable behaviour that occurs in most deterministic, nonlinear dynamical systems. Many mathematical concepts dating back to the last century — and, until recently, regarded by physicists as irrelevant to the description of natural phenomena — have suddenly become important. Notions such as the 'Hausdorff dimension' and 'fractals' have replaced Euclid's dimensions and straight lines, because schoolbook geometry is of little use when confronted with the bewildering complexity generated by nonlinear dynamical systems. The work of B.B. Mandelbrot and especially his book *The Fractal Geometry of Nature* (Freeman, New York, 1982) has been instrumental in turning physicists' attention to

such fractal structures and introducing the relevant mathematical concepts.

A decade ago, M.J. Feigenbaum (*J. stat. Phys.* **21**, 669–706; 1979) conjectured that many physical systems should make the transition from regular (periodic) to chaotic motion in a qualitatively and quantitatively universal fashion, through sequences of period doublings. His approach was very unconventional: instead of modelling a realistic dynamical system he used a computer to iterate a simple map. The regime where chaos occurs is quite inaccessible by standard approximation techniques; but the map reduces the problem to its essence, making it possible to follow the strongly nonlinear behaviour with a minimum of effort and still extract universal quantities. The theory has been confirmed experimentally in many physical systems, ranging from convective flows in liquids to cardiac arrhythmias in chicken hearts. Original articles are collected in two reprint collections: Hao Bai-Lin (ed.), *Chaos* (World Scientific, Singapore, 1984); and *Universality in Chaos* (Hilger, Bristol, 1984; edited by P.C.).

In their experiment, Gwinn and Westervelt apply a time-varying voltage to a crystal of p-type germanium and record the current passing through it. The system has an oscillatory instability so that even a static (d.c.) voltage larger than some threshold value induces a periodic oscillation of the current, with an extremely

stable frequency. When the additional time-varying component is added to the voltage, the two frequencies compete, and if the amplitude of the driving frequency is large enough, chaotic behaviour develops. The interaction of pairs of frequencies is of considerable theoretical interest because of the generality of the phenomenon. As the energy input into a dissipative dynamical system is increased, typically first one and then another intrinsic mode of the system is excited, and their interaction can give rise to chaos. Competing modes usually give rise to mode-lockings: the frequencies adjust slightly to fall into step, making their ratio a rational number. In that case the response is periodic; an irrational ratio corresponds to quasiperiodic behaviour — the motion never quite repeats itself.

The aim of the experiment was to test the theory of the transition from quasiperiodicity to chaos. Mode-lockings (which have interesting properties of their own) are avoided by tuning the external frequency so that its ratio to the internal frequency equalled the golden mean, $(\sqrt{5}-1)/2$. The choice of this ratio does not represent a return to mediaeval alchemy, but is dictated by number theory: the golden mean belongs to a family of irrational numbers for which it is hardest to give good rational approximations. As experimental measurements have limited accuracy, physicists usually do not expect number-theoretic subtleties, such as how irrational a number is, to be of any physical interest. In the theory of transitions to chaos, however, the starting point is the enumeration of asymptotic motions of a dynamical system, and it is through this enumeration that number theory enters and comes to have a central role.

The output of the experiment is the time variation of the current, the full record of which contains much superfluous

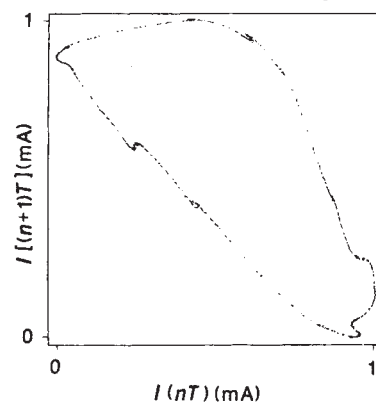


Fig. 1 The strange attractor at the onset of chaos. The current (a constant current of 5 mA has been deducted) is plotted for each cycle of the external drive. T and n , respectively, are the period and number of drive cycles and the current at cycle $n+1$ is plotted against the value at cycle n . (From Gwinn, E.G. & Westervelt, R.M. *Phys. Rev. Lett.* **59**, 157–160; 1987.)

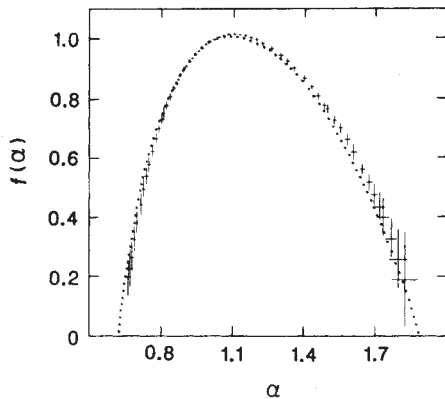


Fig. 2 The function of $f(\alpha)$ quantifying how often a given scaling index or pointwise dimension occurs on the attractor. The error bars indicate the standard deviation of the mean of three data sets. The universal circle-map prediction is shown as a dotted curve. (From Gwinn, E.G. & Westervelt, R.M. *Phys. Rev. Lett.* **59**, 157–160; 1987.)

information. Previously, the results would have been presented as a frequency power spectrum. Today the data (50,000 data points would be typical) are manipulated in the way that a theorist analyses a numerical simulation. First, the system is reduced from a continuous time recording to a discrete series of stroboscopic flashes. This method, devised a century ago by H. Poincaré, makes it possible to survey the dynamics visually. The experimental Poincaré map is shown in Fig. 1. The value of the current at each period of the external drive is plotted against its value at the next period. (The first hundred or so cycles, the transients, are not plotted.) The fact that the response is quasiperiodic implies that the dots on the figure would eventually fill up some closed curve. If the transients had been shown on Fig. 1, we would see points initially far away rapidly approaching that curve, which is therefore an 'attractor'. Precisely at the onset of chaos — which was reached in the experiment by varying the amplitude of the drive — the kinky structure visible on the figure emerges. It is called a 'strange attractor' because of the unusual way points are distributed on it and this distribution contains information about universal features of the transition to chaos.

The strangeness of the attractor can be probed by looking at the distribution of points around some reference point. For a reference point, P , on the attractor we define $N_p(r)$ as the number of neighbours within distance r along the attractor. On a usual (non-strange) quasiperiodic attractor, the points are distributed smoothly; a segment of the attractor looks like a line. Thus, $N_p(r)$ scales with r as $N_p(r) \propto r$ for any point P . For the strange attractor in the figure this is not necessarily true. If we pick a point at random, the theory predicts $N_p(r) \propto r^\alpha$ with α varying between 0.6326... and 1.8980..., and the point P is said to have the

pointwise dimension α .

A powerful formalism for confronting the experiment with the predictions of the theory was developed recently by T.C. Halsey *et al.* (*Phys. Rev. A* **33**, 1141–1151; 1986). They look at the attractor as a superposition of 'subfractals'; namely, for each α , the set of points with pointwise dimension α . A function $f(\alpha)$ is introduced to quantify how often a given value α appears on the attractor — more precisely, $f(\alpha)$ is the Hausdorff dimension of the set of points with pointwise dimension α . Figure 2 shows this curve for the theory and the experiment. The dots are the theoretical prediction computed by iterating a so-called circle-map. The crosses represent $f(\alpha)$ calculated from the experimental attractor shown in Fig. 1. Given that there are no adjustable parameters, the agreement between the theory and the experiment is remarkable. (The

error bars at the edges of the α -interval come from the sparseness of points contributing in those limits; here the finiteness of the data sets and noise play a large part.)

Theory and experiment thus interact in a fruitful way, and it is fascinating that the fractal properties of strange attractors can be measured with such precision in physical systems. Theoretically the most striking fact, not anticipated until a decade ago, is that the dynamical systems approaching chaos do so in a universal fashion. Other fields of physics, notably those concerned with growth or aggregation, reveal all kinds of fractal behaviour, but it is not yet clear what kind of universality (if any) to expect there. □

Tomas Bohr and Predrag Cvitanović are at the Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen, Denmark.

Intermediate filaments

Looking for a function

Benjamin Geiger

STUDIES on the molecular properties of the cytoskeleton are largely motivated by the desire to understand the functions of these cytoplasmic filaments in cells and tissues. Physiological studies, morphological observations and biochemical characterization of proteins in the cytoskeleton have allowed molecular models, albeit tentative, of the possible cellular activities in which the various classes of filaments take part. These models, despite their tendency to be oversimplified, have contributed a great deal to current concepts of mechanisms of cell motility, mitosis, transcellular transport, adhesion, modulation of membrane activity and cellular morphogenesis. Now, however, experimental work^{1–4} is providing a test for many of these ideas about the physiological functions of intermediate filaments, one class of cytoskeletal filaments.

Until recently, most information has come from the other two classes of cytoskeletal filaments, microfilaments and microtubules, which, together with batteries of associated proteins, have been extensively characterized. Moreover, the availability of excellent and well-studied model systems such as the contractile unit of skeletal muscle or the interaction of dynein with microtubules in cilia and flagellae have provided clues about the cytoplasmic activities in which actin and tubulin are involved.

The structure–function relationships of the third cytoskeletal network, the intermediate filaments, are less well characterized. Despite the fact that the primary structure of many intermediate-filament subunits is known and their cellular

distribution extensively documented, only limited molecular information has so far been available on their behaviour *in vivo*. In the absence of specific data, it has been suggested that intermediate filaments are involved in mechanical integration of cytoplasmic space⁵ or in a skeletal framework of the cytomatrix (see refs. 6,7 for reviews). Well-controlled experiments relating their molecular properties to specific cytoplasmic events are, however, still needed to understand their functions in more detail.

Inagaki *et al.*¹ recently investigated specific phosphorylation events in modulating the assembly of vimentin molecules into intermediate filaments. These authors show that vimentin is an excellent *in vitro* substrate for protein kinase C and cyclic AMP-dependent protein kinase, but not of several other kinases. Moreover, phosphorylation by the cyclic-AMP-dependent kinase induces the dramatic disassembly of vimentin filaments. Analysis of tryptic phosphopeptide maps indicates that the sites of phosphorylation with the two kinases are distinct, and the authors suggest that a single, site-specific phosphorylation of vimentin (as well as its

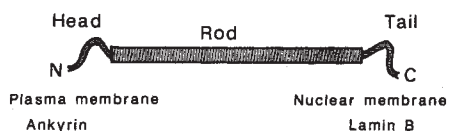


Fig. 1 Intermediate-filament subunit (vimentin or desmin) presenting binding sites for plasma-membrane-associated ankyrin at the amino-terminal head domain, and binding sites for lamin B at the carboxy-terminal tail region.

dephosphorylation) is important in the modulation of filament structure and organization.

These new findings could be related to earlier observations⁸⁻¹⁰ which correlated changes in intermediate-filament structure in cells with phosphorylation events. The phosphorylation-dependent disassembly of intermediate filaments may produce a molecular solution for a cellular dilemma: how do dynamic cellular processes such as cytokinesis or changes in cell shape occur in the presence of the notoriously insoluble network of intermediate filaments? Data obtained using the immunofluorescence and electron microscopes indicate that cells can transiently dissociate their intermediate filaments and pack their subunits into non-filamentous spheroidal structures¹¹⁻¹³. It is not clear whether a site-specific phosphorylation-dependent dissociation similar to the one documented by Inagaki *et al.*¹ is involved, or whether other modifications are effective during the dynamic rearrangement of the cytoplasm *in vivo*. To elucidate this question, it is necessary to define and localize the critical phosphorylation site along the vimentin molecule and to determine directly if specific, spatially controlled phosphorylation-dephosphorylation events are indeed involved in the physiological modulation of filament integrity and structure.

Another related matter concerns the native interactions of vimentin or other intermediate filaments with various cytoplasmic organelles. Many previous studies, based on immunocytochemistry and localization in the electron microscope, pointed to an apparent association of intermediate filaments with many cellular structures. It had been claimed that in addition to their well-documented interactions with desmosomes, intermediate filaments are often detected in close association with the nucleus, the plasma membrane, dense bodies of smooth muscle, microtubules, fat globules of adipocytes and other elements (see ref. 6 for a review). These observations stimulated much speculation concerning the roles of intermediate filaments as specific cytoplasmic organizers, yet the basis for the presumptive interactions remained obscure.

In a series of recent biochemical studies, Georgatos and Blobel^{2,3} show that purified vimentin binds to different fractions of avian erythrocyte membranes through two distinct domains. Sites located at the carboxy-terminal tail of the vimentin molecule bind specifically to nuclear envelopes in a cooperative fashion, whereas the plasma-membrane fraction interacts in a saturable manner with the amino-terminal head of the vimentin molecule. The central rod segment of vimentin has no apparent affinity to either fraction (see Fig. 1). Binding studies reveal specific vimentin-binding constituents along these two types of membrane.

It is shown that major nuclear binding sites for vimentin are lamin B or the lamin A-lamin B hetero-oligomers. Interestingly, lamins A and C, which associate with lamin B to form the nuclear lamina at the endofacial surfaces of the nuclear membrane, have a similar primary and secondary structure to intermediate filaments^{14,15}. Other new work⁴ shows that the membrane-skeleton protein ankyrin could provide a second type of binding site to the amino-terminal head region of vimentin or of desmin, a related intermediate-filament protein of muscle cells (Fig. 1).

The specific interaction of vimentin, and possibly other intermediate-filament subunits with lamin B, raises an interesting

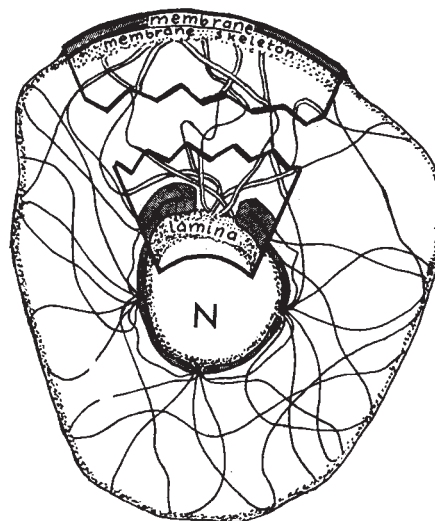


Fig. 2 Hypothetical picture of the intermediate-filament network in cells. Based on the results described in the text, vimentin- or desmin-containing filaments associate with the cell nucleus, interacting with the nuclear lamina through the nuclear pores. At the cell periphery the same intermediate filaments apparently associate with the membrane skeleton, thus forming an elaborate system of nucleolemmal/plasmalemmal interactions. The subunits carrying the two types of binding sites are distributed in a non-polar fashion along the filament.

problem. In their heteropolymeric form, lamins are known to be confined to the endofacial surfaces of the nuclear envelope, whereas intermediate filaments are located throughout the cytoplasm and outside the nucleus. Thus, physical interaction between vimentin and lamins (if and when it takes place *in vivo*) could occur through the nuclear pores, where elements of the nuclear lamina are exposed to the cytoplasm. This suggestion of a direct and specific interaction between elements of the cytoplasmic matrix and peripheral elements of the nuclear matrix is not new. Ultrastructural studies carried out several years ago suggested that such interactions exist and take place through the nuclear-pore complexes¹⁶⁻²⁰. The identification of another binding system along the plasma membrane provides a possible mechanism

for nucleolemmal/plasmalemmal anchorage and communication (see scheme in Fig. 2). The new *in vitro* results¹⁻⁴ seem to complement the existing data concerning the molecular basis and possible mechanisms underlying the physiological interactions of intermediate filaments.

It is too early to evaluate the functional significance of these new observations, yet specific questions and working hypotheses can now be put forward relating the new molecular and structural findings to specific cytoplasmic and nuclear events. Does the nuclear lamina that is exposed at the nuclear pores act as an organizing centre for intermediate-filament assembly? Is ankyrin (or a homologous molecule in non-erythroid cells) a physiological linker of intermediate filaments to the plasma membrane? Are the interactions of vimentin with the nucleus or with the plasma membrane modulated by post-translational modifications of either of the molecules involved? Do the mechanical interconnections between the nucleus and the membrane play a general role in the spatial organization of the cytoplasmic matrix or in the active or passive transport of macromolecules between the nucleoplasmic and cytoplasmic compartments? Are there other intermediate filament-associated proteins on different cellular organelles? How are the site-directed phosphorylations (or dephosphorylations) of vimentin spatially and temporally regulated? Do other intermediate filaments, such as keratins, neurofilaments or glial filaments, have binding properties similar to those of vimentin and desmin, and are they comparably affected by specific phosphorylations? The tools are now ready for concerted efforts to answer these and other questions about the organization and dynamics of the cytoplasm. □

1. Inagaki, M., Nishi, Y., Nishizawa, K., Malsugama, M. & Sato, C. *Nature* **328**, 649-652 (1987).
2. Georgatos, S.D. & Blobel, G. *J. Cell Biol.* **105**, 105 (1987).
3. Georgatos, S.D. & Blobel, G. *J. Cell Biol.* **105**, 117 (1987).
4. Georgatos, S.D., Weber, K., Geisler, N. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6780 (1987).
5. Lazarides, E. *Nature* **283**, 249 (1980).
6. Traub, P. *Intermediate filaments* (Springer, Berlin, 1985).
7. Wang, E., Fishmann, D., Liem, R. & Sun, T.T. (eds) *Ann. N.Y. Acad. Sci.* **445** (1985).
8. Gard, D.L. & Lazarides, E. *Molec. Cell Biol.* **2**, 1104 (1982).
9. Evans, R.M. & Fink, L.M. *Cell* **29**, 43 (1982).
10. Celis, J.E., Larsen, P.M., Fey, S.J. & Celis, A. *J. Cell Biol.* **97**, 1429 (1983).
11. Horwitz, B., Kupfer, H., Eshhar, Z. & Geiger, B. *Expl. Cell Res.* **134**, 281 (1981).
12. Franke, W.W., Schmid, E., Grund, C. & Geiger, B. *Cell* **30**, 103 (1982).
13. Lane, E.B., Goodman, S.L. & Trejdosiewicz, C.K. *EMBO J.* **1**, 1365 (1982).
14. McKeon, F. *et al. Nature* **319**, 463 (1986).
15. Fisher, D.Z., Cahnthy, N. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6450 (1986).
16. Franke, W.W. *Protoplasma* **73**, 263 (1971).
17. Jones, J.C.R., Goldman, A.E., Steinert, M., Yuspa, S. & Godman, R.D. *Cell Motil.* **2**, 197 (1982).
18. Capco, D.G. & Penman, S. *J. Cell Biol.* **96**, 896 (1983).
19. Goldman, R. *et al. Ann. N.Y. Acad. Sci.* **455**, 1 (1985).
20. Lehto, V.-P. *et al. Nature* **272**, 175 (1978).

Benjamin Geiger is in the Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel.

High-energy physics

Mixed beauty at Hamburg

David J. Miller

AMONG the many exciting developments in particle physics reported at a recent meeting*, the most interesting came from the host organization, DESY (Deutsches Elektronen-Synchrotron). The ARGUS collaboration there has seen positive evidence of the 'mixing' of a meson, containing a beauty quark, with its antiparticle. ARGUS has also found a surprisingly large rate for the production of baryons containing beauty quarks. The standard model of particle physics survives intact for another year, having passed several new tests with no serious contradiction of its ingredients: electroweak theory, quantum chromodynamics and the number of quark 'generations'. S.L. Glashow (Harvard) suggested three alternative "fables" of what might emerge at the next generation of accelerators.

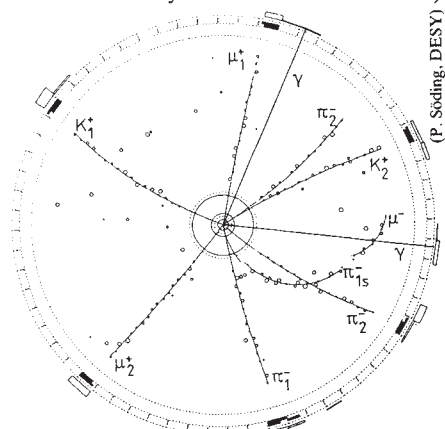
Particle-antiparticle mixing has been studied for more than 20 years with K^0 mesons. These particles consist of a down quark and a strange antiquark, ($d\bar{s}$), but when they decay they sometimes appear to have changed into the $\bar{K}^0$ antiparticle ($\bar{d}s$). In their decays they behave like two orthogonal linear combinations of K^0 and $\bar{K}^0$, called K-long and K-short because of their different lifetimes. An early success of the standard model was to explain K^0 - $\bar{K}^0$ mixing at the same time as predicting the charmed quark. Now the ARGUS group (W. Schmidt-Parzefall, DESY) has seen three distinct independent signals for B^0 - $\bar{B}^0$ mixing ($\bar{B}^0$ is a meson with down antiquark and beauty quark ($\bar{d}b$); B^0 is db).

In both B^0 and K^0 decays, the identity of the decaying quark (particle or antiparticle) is most clearly defined when it decays to a charged lepton (either an electron or a muon), accompanied by a lighter quark and a neutrino (or antineutrino). This is similar to nuclear β -decay, in which an electron (e^-) and an antineutrino are produced when a down quark in a neutron (udd) turns into an up quark in a proton (uud). Down, strange and beauty quarks, all being negative, decay to give a negative electron or muon (μ^-), but their antiquarks give a positive lepton (e^+ , μ^+).

The ARGUS group produced pairs of B^0 and $\bar{B}^0$ in e^+e^- collisions at an energy where no extra particles can be produced. In the absence of mixing, one would expect the decay particles to include pairs of leptons of unlike-sign (one from the particle, the other from the antiparticle). But in one sample the group saw 5 like-sign lepton pairs and 23 with unlike-sign. The group also achieved a rare success: a

fully reconstructed event in which enough tracks could be observed to give a unique identification (see figure). Both decaying particles were B^0 , though one of them must have been produced as $\bar{B}^0$. The UA1 collaboration at CERN (quoted by Schmidt-Parzefall) also reports a like-sign dimuon signal that can be interpreted as B^0 - $\bar{B}^0$ mixing, but not as clearly.

These results put tight constraints on some of the arbitrary parameters of the standard model; in particular on the parameters of the unitary matrix (the Cabibbo-Kobayashi-Maskawa matrix)



The completely reconstructed event from the ARGUS experiment in which B^0 - $\bar{B}^0$ mixing was identified. A B^0 - $\bar{B}^0$ pair is created by an electron-positron collision. The antiparticle turns into a particle by mixing. The two particles then undergo a series of decays to emit the detected particles μ^+ , π^+ , π^- , K^+ in the first case and μ^- , π^- , π^+ , K^- and two photons γ in the second. The open circles mark the detected track and the solid lines the computed track.

that describes the weak couplings of all the quarks. The matrix can still be kept unitary (M. Schifman, Institute of Theoretical and Experimental Physics, Moscow; A. Sirlin, New York University), which is necessary if the theory is to conserve the overall sum of quarks and antiquarks, but only if another free parameter of the model is constrained: the mass of the top quark would have to be at least 50 GeV, contradicting the CERN claim (now retracted) to have discovered top between 30 and 50 GeV at UA1 (see *Nature* 311, 210; 1984). This puts the threshold for $t\bar{t}$ production out of range of the Stanford Linear Collider now tuning up, but within range of Large Electron-Positron project (LEP) phase 2 at CERN.

Researchers at CERN have made an improved measurement of the violation of 'CP' symmetry, in which the behaviour of a particle is indistinguishable from that of its antiparticle, as seen in a mirror. The

spontaneous breaking of this symmetry (first observed in K-meson decays in 1964) may be a clue to the physics of the cosmological Big Bang (reviewed by M.S. Turner, Chicago University). The parameter in the Cabibbo-Kobayashi-Maskawa matrix that describes CP violation should cause a small difference between the decay rates of K-short mesons to two charged or two neutral pions; a difference that was confirmed for the first time in the CERN experiment (I. Mannelli, University of Pisa).

In principle, CP violation should be observed in B-meson decays, and its rate can be predicted as the mixing parameters are known. Present accelerators, however, would take 100 years to supply the data to measure the effect (Schmidt-Parzefall) and a new generation of medium-energy, high-luminosity, electron-positron colliders is proposed to complement the high-energy ones being constructed now (described recently in *News and Views: Nature* 329, 107; 1987).

The newest result reported at the meeting is also the hardest to fit into the standard model (Schifman). The ARGUS collaboration has measured an unexpectedly high rate for the production of pairs of beauty baryon and beauty antibaryon. (Beauty baryons contain a beauty quark and two light quarks; bud, for example.) But there is much work to be done, both experimental and theoretical, before this can be taken seriously. Schifman spent some time explaining how the intrinsic complexity of the "kitchen of strong interactions" destroys some of the simple predictions of pure electroweak theory, and may affect beauty baryon production.

Oddly enough, a similar concept of Schifman's kitchen, but called 'weather' (also denoting nonlinear fluctuations at an unobservably small scale) was introduced by D. Schramm (University of Chicago) to explain disagreements between the standard model of supernovae and the optical observations of the recent supernova 1987A. The neutrino signals are thought to come through the weather like radar through fog, and agree very well with models of the simple primary process. Glashow's three fables, each a minimal alteration of the standard model within present experimental constraints, were presented as examples of the new physics (heavy decaying neutrinos; extra Higgs particles; 'chiral colour' giving a new particle zoo) that could emerge at LEP, or the proposed Superconducting Supercollider in the United States. Although explaining some features of the standard model, the fables avoid any reference to such ambitious (and experimentally unsupported) schemes as technicolour, supersymmetry or superstrings. □

David J. Miller is in the Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.

* International Symposium on Lepton and Photon Interactions at High Energy, 27-31 July 1987, Hamburg, FRG

Neurobiology

Taking apart NMDA receptors

Alan C. Foster and Graham E. Fagg

If the frequency of publications in *Nature* is an indication of the frontiers of scientific research, then excitatory amino-acid receptors, and specifically, the NMDA (*N*-methyl-D-aspartate) receptor, must rank as one of the 'hottest' topics in neurobiology at the moment. Why is this? In a large part, it is because the NMDA receptor — one of the classes of glutamate receptor mediating excitatory synaptic transmission — is being found to play a critical role in many complex neurophysiological and 'plastic' events which underlie central neural development and function in vertebrates. Recent work shows that NMDA receptors are involved in long-term potentiation, burst firing and the generation of patterned activity in neuronal networks; in use-dependent stabilization of synaptic connectivity in young animals; in some forms of learning in mature animals; and in pathologies such as epilepsy, spasticity and ischaemic neuronal death (see refs 1, 2 for reviews).

What do we know about the molecular organization and structure of the NMDA receptor? Can its contribution to brain physiology and pathology be rationalized in terms of the receptor channel complex? Although precise correlations of structure and function await the application of molecular-biological techniques, data now emerging from electrophysiological and biochemical investigations show that the NMDA receptor comprises distinct domains (see figure) whose interactions form the basis of the initiation and regulation of receptor activity. The NMDA receptor complex is not simply a ligand-gated ion channel (as are many fast-acting transmitter receptors), but is

subject to additional regulation by the simple amino acid glycine and, uniquely, by voltage-dependent binding of magnesium ions inside the channel. Yet another binding site for psychoactive drugs such as phencyclidine (PCP) and MK-801 is also located within the channel.

What is known about the domains in the NMDA receptor complex? The transmitter recognition site is the most extensively characterized pharmacologically. Here, L-glutamate and L-aspartate bind with high affinity, whereas analogues of dicarboxylic amino acid such as AP5 (2-amino-5-phosphonopentanoate) and CPP (3-[2-carboxypiperazin-4-yl] propyl-1-phosphonate) are the most potent and widely used competitive antagonists³. There are hints from recent structure-activity studies, however, that such antagonists bind partly to another attachment point close to the transmitter-recognition site and that the charge at the ω -terminal is a determinant of agonist versus antagonist properties⁴.

The most significant discovery for understanding the channel domain of the NMDA receptor was that the agonist-evoked conductance is blocked in a voltage-dependent manner by magnesium ions^{5,6}. Single-channel experiments show that this unusual form of voltage-gating largely results from blockade of the open channel by Mg^{2+} , probably at a site deep in the channel, near to the intracellular face of the membrane. At membrane potentials close to resting and at normal extracellular Mg^{2+} concentrations (1–2 mM), the channel is blocked and agonists induce little current flow. As the membrane is

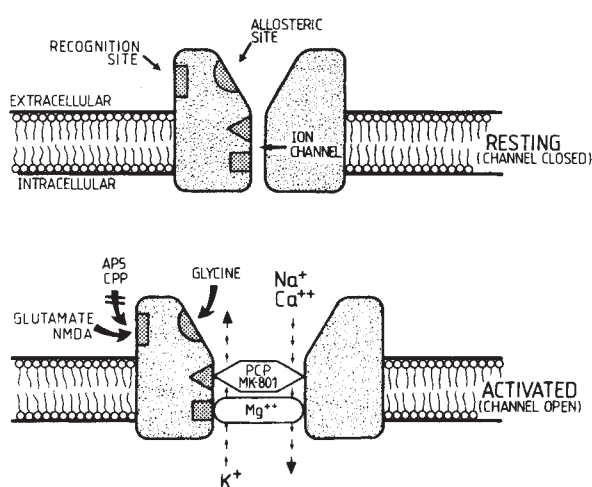
depolarized, the block is relieved, probably because the decreased transmembrane electric field less effectively 'pulls' Mg^{2+} into the channel. The result of this voltage-dependent channel block is that the NMDA receptor conductance exhibits regenerative properties which, together with Ca^{2+} - or depolarization-triggered hyperpolarizations, underline its involvement in phenomena such as rhythmic firing and epileptiform activity.

Recently, several organic molecules (such as PCP, ketamine, MK-801), all of which interact with 'PCP-binding sites' in brain membranes, have been shown to antagonize non-competitively NMDA receptor responses. The blocking actions of these substances are dependent on membrane voltage (like Mg^{2+}) and on activation of the receptor with an agonist such as NMDA (use-dependent)^{8,9}. Such use-dependence is also apparent in radioligand binding experiments⁹. Very recent work shows that recovery from blockade by ketamine is also dependent on agonist application, suggesting that this substance can become 'trapped' inside the channel, and an analysis of the voltage-dependence of the block indicates that ketamine interacts with a site roughly midway across the membrane field¹⁰. These data are consistent with the hypothesis that PCP and related compounds block NMDA receptor responses by binding inside the open channel and impeding transmembrane ion fluxes.

Do PCP and MK-801 interact with the same site as magnesium ions? Divalent cations interact with the binding site labelled by [³H]thienyl-PCP (TCP) and [³H]MK-801 in a complex manner, causing both enhancement and inhibition of binding at appropriate concentrations^{11,12}. Whereas all divalent cations tested to date inhibit binding, only those possessing NMDA-antagonist properties enhance binding, suggesting that this latter effect represents an interaction within the NMDA-receptor ion channel and that PCP/MK-801 and Mg^{2+} bind to separate sites within this domain.

The discovery of a third domain within the NMDA receptor complex, through which glycine regulates receptor activity, began with observations that this simple amino acid markedly augments NMDA-evoked responses in cultured neurons¹³. This effect is strychnine-insensitive, and results principally from an increase in the frequency of NMDA-elicited channel opening. The possibility that glycine is a specific regulator of NMDA-receptor mechanisms has rapidly gained credence from the similar regional distributions of strychnine-insensitive glycine binding sites and NMDA receptors in the brain¹² and, more directly, from very recent findings that the stimulation of [³H]TCP and [³H]MK-801 binding by NMDA-receptor agonists is potentiated by this

Model of the NMDA receptor channel complex. Top, closed; bottom, open. The receptor appears to be proteinaceous⁴ and may consist of multiple subunits. Four sites in the complex have been identified on the basis of their function: (1) the transmitter recognition site, with which agonists (NMDA, L-glutamate) and competitive antagonists (AP5, CPP) interact; (2) a site through which glycine allosterically up-regulates receptor function, possibly by facilitating agonist-induced transitions to an open-channel configuration; (3) a cation-binding site inside the channel, where (depending on membrane potential) Mg^{2+} can bind and block transmembrane ion fluxes; and (4) a PCP-binding site, which requires agonist activation at the transmitter recognition site, interacts with the divalent cation binding site, and at which psychoactive drugs such as PCP and MK-801 bind to function as open-channel blockers.



amino acid^{4,11,15}. Although the details of the interactions between receptor domains remain to be elucidated, these data suggest the fascinating hypothesis that, in contrast to its inhibitory function in spinal tissue, glycine acts in higher brain areas to facilitate allosterically NMDA receptor-mediated excitation.

The domain structure of the NMDA receptor (see figure) is, in many respects, similar to that which has been described for other transmitter-gated ion channels. At both the nicotinic acetylcholine and GABA_A receptors, 'non-competitive' antagonists interact with the receptor complex in an agonist-dependent manner, reflecting, as in the case of PCP or MK-801 at the NMDA receptor, a preference for activated or open states of the ion channel. At the GABA_A receptor, the well-documented ability of benzodiazepines to enhance receptor activity allosterically can be compared with the action of glycine at the NMDA receptor. As well as these common features, the NMDA receptor has evolved a unique form of voltage-dependent gating, involving binding of Mg²⁺ inside the channel. Hence, 'switching on' NMDA receptor mechanisms requires at the minimum: (1) binding of transmitter to its recognition site; and (2) relief of the channel blockade (for example, by local depolarization of the postsynaptic membrane by other excitatory inputs). In addition, receptor function may be subject to rapid up- or down-regulation by changes in the levels of glycine and/or Mg²⁺ in the local extracellular environment.

Now that the basic functional organization of the NMDA receptor is understood, what are the next steps? First, understanding the full physiological implications and intricacies of its regulatory mechanisms requires detailed studies of the interactions between receptor domains and their stoichiometric relationships, the modulation of different affinity states, and the seemingly diverse effects of divalent cations. In view of recent reports that Zn²⁺ non-competitively antagonizes NMDA-evoked responses at a site distinct from that at which Mg²⁺ binds^{16,17}, one might ask how many more domains of potential significance there are. Second, although the biophysical properties of the NMDA receptor-channel complex can account for its involvement in electrical phenomena such as signal amplification and rhythmic firing, they cannot alone explain its role in prolonged events like long-term potentiation or synaptic reorganization. In this respect, a clue for future studies is that the NMDA receptor channel is permeable to Ca²⁺ (ref. 18), which may activate long-term biochemical changes in the post-synaptic neuron.

The goal of the reductionist approach is to understand function in terms of molecular structure. Such studies are

John Howard Northrop (1891–1987)

JOHN NORTHROP probably did more than any other individual to establish the view that pure enzymes are indeed proteins. He was not, however, the first to crystallize an enzyme. That honour belonged to J.B. Sumner of Cornell University, who reported the crystallization of urease from jack bean in 1926. Northrop's work, however, was both more extensive and more searchingly critical than Sumner's. Both men had to face widespread scepticism concerning their work: Richard Willstätter in Munich, with his great prestige, had denied the protein nature of enzymes. Most biochemists, particularly in Europe, followed Willstätter, and for years discounted the evidence of Sumner and Northrop.

Northrop, born in Yonkers, New York, did both his undergraduate and graduate work at Columbia University, receiving his PhD in 1915. In 1916 he became an assistant in the laboratory of Jacques Loeb at the Rockefeller Institute in New York. Loeb was a major influence in his development. From the beginning Northrop worked on enzymes; many of his early papers concerned the kinetics of enzyme-catalysed reactions. After Loeb's death in 1924, Northrop arranged to be transferred to the Princeton branch of the Rockefeller Institute. He loved open country, and strongly disliked living in the city.

It was in Princeton that he achieved the crystallization of three major proteolytic enzymes — pepsin in 1930, and trypsin and chymotrypsin a few years later — and provided the most rigorous test then available to demonstrate that they are pure proteins. He well knew that the preparation of a substance in crystalline form is not in itself an adequate criterion of purity; mixed crystals of two or more closely related substances were a common occurrence. Northrop's test for purity involved careful solubility studies in well-defined media, with application of the phase rule of Willard Gibbs. A truly pure protein in such a medium should give a constant solubility, independent of the amount of crystalline

phase in equilibrium with the solution. He was able to show, for his preparations, that this was indeed the case. Sumner, in his pioneer work, had not applied such rigorous tests; so on the whole it was Northrop's work that was most influential in gradually convincing the sceptics.

Northrop's associate, Moses Kunitz, shared much of this work, and later continued independently to crystallize other enzymes, notably ribonucleases and deoxyribonucleases. Northrop and Kunitz, together with Roger Herriott, described this research in their book *Crystalline Enzymes*.

In 1938, the trustees of the Rockefeller Institute closed the Princeton branch of the Institute. Most of the Princeton workers moved back to New York or retired. Northrop, however, refused to live in New York and the trustees permitted him to move elsewhere, to a place of his choice, while still remaining a member of the Institute and receiving its support for his work. In 1939 he settled at the University of California at Berkeley, where he became a professor and did extensive work on bacteriophage, his approach being similar to that of his earlier work on enzymes. But he missed the biological significance of the life cycle of bacteriophage, and his findings were eclipsed by the work of molecular biologists.

In 1946, Sumner and Northrop received the Nobel prize in chemistry (together with Wendell Stanley, for his work on viruses). At last the sceptics had been convinced of the validity of Northrop and Sumner's work. This was the chief of the many honours that Northrop received. He retired from Berkeley in 1959, and characteristically chose to live far away from noise and bustle, in the rather remote district of Wickenburg, Arizona, where he died.

John T. Edsall

John T. Edsall is in the Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

under way for the nicotinic acetylcholine receptor¹⁹ and, given its unique properties, these will undoubtedly be of special interest for the NMDA-receptor complex. The path ahead is certainly long, but the recent work described here is a valuable initial step towards elucidating the molecular basis of some of the most intriguing excitatory and plastic events in the central nervous system. □

1. Fagg, G.E. *et al.* *Trends pharmac. Sci.* **7**, 357–363 (1986).
2. Cotman, C.W. *et al.* *Trends Neurosci.* **10**, 263–302 (1987).
3. Watkins, J.C. & Olverman, H.J. *Trends Neurosci.* **10**, 265–272 (1987).
4. Fagg, G.E. & Baud, J., in *Excitatory Amino Acids in Health and Disease* (eds Lodge, D. *et al.*) (Wiley, Chichester, in the press).
5. Nowak, L. *et al.* *Nature* **307**, 462–465 (1984).
6. Mayer, M.L. & Westbrook, G.L. *Prog. Neurobiol.* **28**, 197–276 (1987).
7. Lodge, D. *et al.* in *Excitatory Amino Acid Transmission* (eds Hicks, T.P. *et al.*) 83–90 (Liss, New York, 1987).
8. Honey, C.R. *et al.* *Neurosci. Lett.* **61**, 135–139 (1985).
9. Kemp, J.A. *et al.* *Trends Neurosci.* **10**, 294–298 (1987).
10. MacDonald, J.F. *et al.* *J. Neurophysiol.* **58**, 251–266 (1987).
11. Johnson, K.M. *et al.* in *Sigma Opioid Phenylcyclidine-like Compounds as Molecular Probes in Biology* (eds Domino, E.F. & Kamenka, J.M.) (NPP Books, Ann Arbor, in the press).
12. Wong, E.H.F. *et al.* *J. Neurochem.* (in the press).
13. Johnson, J.W. & Ascher, P. *Nature* **325**, 529–531 (1987).
14. Bowers, N.G. *Nature* **326**, 338 (1987).
15. Reynolds, I.J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
16. Peters, S. *et al.* *Science* **236**, 589–593 (1987).
17. Westbrook, G.L. & Mayer, M.L. *Nature* **328**, 640–643 (1987).
18. MacDermott, A.B. *et al.* *Nature* **321**, 519–522 (1986).
19. Mishina, M. *et al.* *Nature* **321**, 406–411 (1986).

Alan C. Foster is at the Neuroscience Research Centre, Merck Sharp and Dohme Ltd, Harlow, Essex CM20 2QR, UK, and Graham E. Fagg is at the Biology Research Laboratories, Ciba-Geigy, 4002 Basel, Switzerland.

When evolution cannot go into reverse

SIR—P.L. Williams (*Nature* 328, 21–22; 1987) incorrectly refers to “ambiguity” in our article on irreversible evolution (*Nature* 326, 128; 1987). For the purpose of our discussion, the process was defined as involving “evolutionary routes that, once taken, preclude return”, and illustrated with an example taken from hymenopterans in which both biology and theory are understood. In none of the cases cited by Williams is there any reason to believe that a strict retracing of evolution could not occur if all ancestral environments were retraced. In the peppered moth example, the complications cited by Williams do not prevent reversible evolution. The whole point of our article was that, on occasion, there are good reasons why an evolutionary route would not be retraced even given the ancestral intermediate environments.

PAUL HARVEY

Department of Zoology,
University of Oxford,
South Parks Road,
Oxford OX1 3PS, UK

LINDA PARTRIDGE

Department of Zoology,
University of Edinburgh,
West Mains Road,
Edinburgh EH9 3JT, UK

Function of a new globin gene

SIR—Recently a new gene ($\theta 1$) has been identified at the 3' end of the α -globin cluster in several species^{1–3} including man. It has been suggested that this gene may be functional in higher primates because in the orang-utan and olive baboon it has all of the structural features necessary for expression and in man it has given rise to a processed pseudogene elsewhere in the genome⁶. Further evidence for its functional status comes from comparisons of the silent versus replacement site substitutions in two primate $\theta 1$ genes which indicate that it has been evolving under selective constraints². Speculation about the function of $\theta 1$ has centred on its possible role as an embryonic globin involved in placentation^{1,2}.

At present no protein or messenger RNA corresponding to the $\theta 1$ gene has been reported. In the past, the function of various globin genes has been established by identifying haemoglobins corresponding to their gene products and observing

the effect of naturally occurring mutants of globin synthesis (thalassaemias) on the phenotype of affected individuals. As far as we know, however, there are no unassigned human globins and therefore no candidate gene products for $\theta 1$. Interestingly, analysis of the predicted amino-acid sequence of the $\theta 1$ counterpart in horse suggests that it may not encode a viable globin-like protein at all (see page 465 of this issue⁷) but as it has clearly been conserved throughout evolution, may subserve some other function.

In man, the $\theta 1$ gene lies at the 3' end of the α -globin complex which includes an embryonic gene (ζ), two adult α genes ($\alpha 2$ and $\alpha 1$) and three pseudogenes ($\psi \zeta 1$, $\psi \alpha 1$ and $\psi \alpha 2$) (see figure). We have recently shown that a determinant for α thalassaemia in South-East Asia results from a deletion that removes both α genes ($\alpha 2$ and $\alpha 1$) and the $\theta 1$ gene (see figure)⁸. Heterozygotes for this determinant have a mild anaemia but are otherwise normal. Homozygotes survive *in utero* until 28–32 weeks of gestation but, with rare exceptions⁹, die in the perinatal period due to the effects of prolonged intrauterine hypoxia; this condition is known as the Hb Bart's hydrops fetalis syndrome. The fact that such infants can survive at all until this late stage in development makes it very unlikely that the $\theta 1$ gene encodes a critical embryonic globin.

Observations on homozygotes for this deletion may also be instructive in evaluating other possible roles for the $\theta 1$ gene. Although infants with the Hb Bart's hydrops fetalis syndrome have a relatively high incidence of congenital malformations, there is no consistently abnormal developmental pattern and it is likely that such abnormalities are secondary to chronic intrauterine hypoxia. We have re-evaluated the genotype of a child with this syndrome who has been intensively treated and kept alive by blood transfusions⁹. We find this child to be homozygous for the South-East Asian α -thalassaemia determinant, both $\theta 1$ genes are deleted (see figure), and yet the child appears to have developed normally, indicating that $\theta 1$ does not encode any other critical protein that cannot be replaced by regular blood transfusion.

It is possible that $\theta 1$ is only one member of a family of θ -like genes¹ and that deletion of a single member ($\theta 1$) is of little or no consequence. However, the observations reported here raise some doubts as to whether $\theta 1$ is expressed as a

protein; it is unlikely to be a part of a viable haemoglobin molecule⁷ and probably does not encode a critical protein, although it may represent a non-essential facultative protein. If this is the case, it raises the interesting question of why the $\theta 1$ gene has been so well conserved throughout mammalian evolution.

N. FISCHER-GHODSIAN

D.R. HIGGS

MRC Molecular Haematology Unit,
Nuffield Department of Clinical Medicine,
John Radcliffe Hospital,
Headington, Oxford OX3 9DU, UK

E.C. BEYER

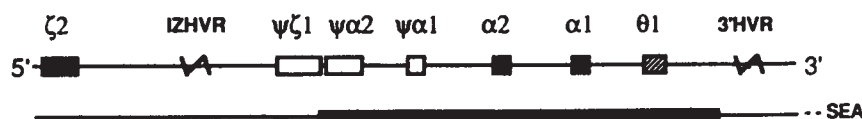
Department of Pediatrics,
Harvard Medical School,
Boston, Massachusetts, USA

1. Marks, J., Shaw, J-P & Shen, C-K.J. *Nature* 321 785–788 (1986).
2. Shaw, J-P., Marks, J. & Shen, C-K.J. *Nature* 326, 717–720 (1987).
3. Cheng, J-F. *et al. J. Biol. Chem.* 261, 839–848 (1986).
4. Sawada, I. & Schmid, C.W. *J. molec. Biol.* 192, 693–709 (1986).
5. Clegg, J.B. *et al. Nucleic Acids Res.* 12, 7847–7858 (1984).
6. Shaw, J-P. *et al. in Hemoglobin Switching V* (eds Stamatoyannopoulos, G. & Nienhuis, A.) (in the press).
7. Clegg, J.B. *et al. Nature* 329, 465 (1987).
8. Nicholls, R.D., Fischer-Ghodsian, N. & Higgs, D.R. *Cell* 49, 369–378 (1987).
9. Bianchi, D.W. *et al. J. Pediatr.* 108, 716–718 (1986).

Th/Nd abundance ratio in the surfaces of G-dwarfs

SIR—In a stimulating recent article, Butcher¹ makes wide-ranging conclusions about the age of the Galaxy from his observations of the essential constancy of the Th/Nd abundance ratio in the surfaces of G-dwarfs of varying age. He argues that in the simplest model of chemical evolution for the Galaxy, one in which the nucleosynthesis rate per unit mass of interstellar medium is constant so that stable-element gas concentrations rise linearly in time, the ratio of the concentration of Th to that of a coproduced stable element observed today in a G-dwarf that had formed at time T rises as $(e^{\lambda T} - 1)/\lambda T \approx 1 + 0.025T$. Accordingly, in terms of the age of the star (galactic age $- T$) the surface concentrations today decrease with increasing age. His data, reproduced in the figure, do not decline with age, leading to Butcher's powerful conclusions.

I write in admiring caution to argue that, from a purely theoretical point of view, the simplest model expectation is of opposite sign and matches the data well (the curve in the figure). In the astrophysically simplest model, the r -process abundances are primary and indeed grow linearly in time, but the s -process abundances are secondary, being proportional to the initial stellar metallicity, so that they grow quadratically in time. Because Th/Nd observed today measures $\text{Th}/(\text{Nd} + \text{Nd}_s)$ at time T , the ratio expressed above should, in terms of the time dependence of this simple model, be

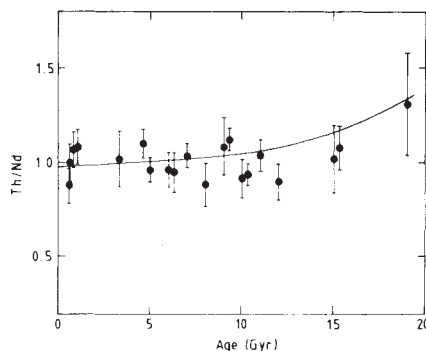


The normal human α -globin complex. The heavy black rectangle below depicts the extent of the South-East Asian deletion.

divided by the additional factor $[(1-s) + (T/T_{\odot})s]$, where $s \equiv \text{Nd}_{\odot}/\text{Nd}_{\odot}$ is the fraction of solar Nd that has resulted from the s -process and $T_{\odot}$ is the solar birthdate. Fortunately the neutron-capture cross sections for Nd isotopes have recently been measured², enabling the conclusion that $s = 0.52$ (slightly more than half s -process). Using this value and $T_{\odot} = 15$ Gyr to match the age distribution of stars in the figure results, with no arbitrary parameters, in the curve shown there as the simplest *a priori* expectation for this concentration ratio. This result differs in two important ways from the monotonic decline used in Butcher's version of the figure. First, the ratio observed today is expected to be essentially constant for stars with ages between zero and about 12 Gyr. Second, the ratio is expected to turn upward for stars older than 15 Gyr. Both features seem to me to match the data of the figure better than the monotonic decline used by Butcher. With these assumptions the expectation would therefore be that the observed Th/Nd ratio should not be useful for galactic chronometry, except for its validation of those assumptions. The distribution in the figure is not strongly dependent upon the assumed galactic age, but the upturn in the oldest stars would be still greater for a younger Galaxy. Clearly, a Galaxy as old as 20 Gyr, as estimated from the stellar ages, is not in conflict with this simple expectation, which makes it more in tune with the old ages derived from the discovery two decades ago of the Re/Os chronology³.

Let it be clear that I in no sense imply that Butcher has erred, for his article shows clearly that he understands and admits this possible loophole in his argument, but that he prefers to believe the implications of his own prior studies⁴, suggesting little variation of the s/r ratio in old stars. However, a constant s/r ratio would, in my view, be astrophysically complicated rather than simple. My own prejudice for theoretical simplicity draws some support from observations^{5,6} that in the oldest metal-poor stars the r -process has occurred prior to the s -process and has accordingly grown faster initially. Thus the spirit of my remarks is not criticism but rather caution — that the profound conclusion that Butcher has given rests ultimately on understanding a major astrophysical puzzle concerning the relative growth rates of s and r abundances with galactic age and upon the conviction that no systematic misunderstanding compromises the difficult interpretation of spectroscopic line strengths that suggest the absence of differential growth⁴.

This caution cannot be easily relaxed by galactic infall. I have presented elsewhere⁷ analytic models of the chemical evolution in the face of time-dependent metal-poor infall in which the radioactive concentrations are analytically expressed⁷, as are



Th/Nd line-strength ratios versus stellar age¹. Solid curve: calculation presented in this work when normalized to fit the young stars. This *a priori* expectation is not in conflict with the old Galaxy required by these stellar ages.

also the secondary metallicities⁸. As long as these models yield monotonic temporal growth of galactic metallicity, they also yield a curve similar to that in the figure.

This work was supported by the Robert A. Welch Foundation, by NASA, and by Durham University. I thank St. Mary's College, Durham, for a research fellowship that supported my residence there during my sabbatical leave from Rice University.

DONALD D. CLAYTON

Department of Physics,
Durham University,
Durham DH1 3LE, UK

1. Butcher, H.R. *Nature* **328**, 127–131 (1987).
2. Mathews, G. & Kappeler, F. *Astrophys. J.* **286**, 810–820 (1984).
3. Clayton, D.D. *Astrophys. J.* **139**, 637–650 (1964).
4. Butcher, H.R. *Astrophys. J.* **199**, 710–717 (1975).
5. Sneden, C. & Pilachowski, C.A. *Astrophys. J.* **288**, L55–L58 (1985).
6. Sneden, C. & Parthasarathy, M. *Astrophys. J.* **267**, 757–778 (1983).
7. Clayton, D.D. in *Nucleosynthesis* (eds Arnett, W.D. & Truran, J.W.) 65–88 (University of Chicago Press, 1985).
8. Clayton, D.D. & Pantelaki, I. *Astrophys. J.* **307**, 441–448 (1986).

BUTCHER REPLIES—As Dr Clayton points out, a judicious choice of s -process abundance evolution over time in the Galaxy can reproduce my data while retaining a large maximum age. In this case, however, one predicts that the ratio of r - to s -process abundances for stable elements will vary in a certain way among stars similar to those in my sample. Variations of the necessary magnitude — a factor of three or more for europium (91% r -process) to barium (84% s -process)¹ — are clearly ruled out by abundance data published elsewhere². This conclusion could only be compromised if the shapes of the abundance curves evolved so that Eu/Ba remained constant while Th/Nd at production changed. I deem the latter unlikely, if not yet completely excluded. My suggestion, therefore, is that undetectable abundance evolution in both radioactive and stable species is evidence for a total timescale much shorter than the relevant decay time, regardless of whether one has an extended production history or not.

The trouble with all these discussions is that the theory of the chemical evolution of the Galaxy by stellar nucleosynthesis has too many unknown parameters, and does not provide much in the way of testable predictions. Therefore, whether a particular development of s - and r -process abundances, with a very large age, is the simplest model must, at this stage, be a matter of taste. Another simple model, for example, which fits available data particularly well except for the stellar-age scale, postulates an initial event with only the most minor ongoing synthesis thereafter. Neither model addresses the curious observation that stars of a given age can vary by factors of five in their heavy-metal-to-hydrogen ratios, while the relative abundances of elements produced by different nuclear processes show undetectable variations. The former suggests mixing of the interstellar gas is not very efficient on short timescales; the latter requires not only efficient mixing, but also a very special production history if synthesis is an ongoing phenomenon.

It seems certain that stars make heavy elements by nuclear reactions in their interior, but how this fact translates into the chemical evolution of the interstellar medium remains vague. My results emphasize that short total timescales and element-production histories not simply tied to the stellar birthrate may well be the preferred possibilities.

HARVEY BUTCHER

Kapteyn Astronomical Institute,
Postbox 800, 9700 AV Groningen,
The Netherlands

1. Seeger, P.A. et al. *Astrophys. J. Suppl. Ser.* **11**, 121–166 (1965).
2. Butcher, H.R. *Astrophys. J.* **199**, 710–717 (1975).

Aluminium and cooking

SIR—Since Savory *et al.*¹ reported far less leaching of aluminium from utensils in the presence of fluoride than we did², we have re-examined our experiments and discovered that our aluminium estimates were in error.

We now agree with Savory *et al.* that there is minimal leaching of aluminium in the presence of 1 p.p.m. fluoride, although at 10 p.p.m. fluoride and above, leaching becomes significant. We have also observed a synergistic effect of chloride. We still believe that fluoride-induced leaching of aluminium may in some circumstances be relevant to the possible cumulative toxicity of aluminium.

K. TENNAKONE

Institute of Fundamental Studies,
Hantana, Kandy, Sri Lanka

S. WICKRAMANAYAKA

Department of Physics,
University of Ruhuna,
Matara, Sri Lanka

1. Savory, J. et al. *Nature* **327**, 107 (1987).
2. Tennakone, K. & Wickramanayaka, S. *Nature* **325**, 202 (1987).

Genetic machinations

Peter Newmark

Invisible Frontiers: The Race to Synthesize a Human Gene. By Stephen S. Hall. *Atlantic Monthly Press, New York: 1987. Pp.334. \$19.95. To be published in Britain next March by Sidgwick & Jackson.*

AT THE end of the summer of 1978, rumours began to spread that one of Britain's top-secret laboratories had been penetrated by foreign agents. This was not the stuff of *Spycatcher*, however, for it was by invitation that four Harvard scientists were spending a month in the Ministry of Defence's Microbiological Research Establishment at Porton while they attempted to clone the human insulin gene.

Two pressures drove them to Porton. The first was that the good folk of Cambridge, Massachusetts, led by their mayor, Alfred Vellucci, and supported by such Harvard scientists as George Wald, were still in the process of holding back our heroes from the frontiers of recombinant DNA research. The result was that no high-containment laboratory had been constructed at Harvard, whereas Porton, with its need to handle highly pathogenic micro-organisms, had such a laboratory available.

The second pressure came from the knowledge that two other groups were also in pursuit of the insulin gene (strictly speaking, it was not the gene itself, but simply the DNA encoding insulin that was being sought). One of the rival groups was centred on the laboratories of Howard Goodman and William Rutter at the University of California, San Francisco. The other was associated with the fledgling biotechnology company, Genentech, whose scientific founder, Herb Boyer, was on the same campus.

As described by Stephen Hall, the four weeks that Wally Gilbert, who had just become involved in a rival company, Biogen, and his three colleagues spent at Porton were pure Gilbert and Sullivan. There were fittings for gas masks, the extreme measures that had to be taken on entering or leaving the high-containment laboratory, and the clash of attitude (and attire) between the regiment of 'gentleman' scientists of the British military establishment and the regimentally casual Harvard men — and women, God forbid! And then contamination of the hard-won human DNA by rat DNA scuppered the purpose of the whole trip.

Worse still, while the Harvard team was struggling in Porton, Genentech announced the production of human insulin from chemically synthesized genes for the hormone's two peptide chains. The use of synthetic genes cleverly avoided the need

for stringent containment facilities when it came to cloning and expressing them. Success came from a collaboration between City of Hope scientists, Arthur Riggs and Keiichi Itakura, who had already synthesized a somatostatin gene with Genentech's backing, and the company's own scientists — still without a laboratory of their own — Dennis Kleid and David Goeddel. The result was a bril-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

End product — Eli Lilly's advertisement for Humulin, genetically engineered human insulin, the outcome of the events described in Hall's book.

liant press for Genentech and several million dollars of backing from Eli Lilly, which was the real stimulus to Genentech's rise.

At the time of Genentech's success, the University of California, San Francisco, was also backed by Eli Lilly. Indeed, one of Goodman's post-doctoral students, Axel Ullrich, was making use of the company's high-containment facility in France. He was there for the same reason that the Harvard team were in Porton and was meeting with as little success. Later, Ullrich and his compatriot Peter Seeburg, who was cloning growth hormone genes, were to break with Goodman and join Genentech.

Stephen Hall is a journalist, and he unravels the various threads in the drama with great skill. His sources are numerous

interviews with all the major actors, apart from Howard Goodman, who declined, and with many of the bit players. We are reminded of the days when molecular biologists lost their innocence and bought suits. When venture capitalists struggled to learn the difference between exon and Exxon. When pharmaceutical companies, by and large, concluded that molecular biology had little to offer them. And when public-interest groups and regulators busied themselves saving the world from being ravaged by man-made bugs.

One incident that sums up much of what was worst about the times is recalled, if not resolved, by Hall. Late in 1976, Ullrich was anxious to begin cloning the rat insulin gene. The obvious way to proceed was to make use of plasmid pBR322, which had been designed for cloning in Boyer's laboratory. The problem was that the regulators at the National Institutes of Health were treading very warily in the face of public suspicion and had not cleared the plasmid for use. In mid-January 1977 Boyer received news that the plasmid had been "approved but not certified". Ullrich, in circumstances that remain uncertain, took this as the signal to proceed with cloning.

Success came quickly. But so did the inescapable realization that the plasmid should not have been used before certification. A series of anguished discussions led to the decision to destroy the clones in mid-March. One month later a plasmid that was almost as good as the pBR322 was certified and, with astonishing speed, Ullrich re-cloned the rat insulin gene. The result was announced before its publication at a full-scale press conference which, apart from being a wild success, taught Ullrich and others that postdocs often toil for the glory of their laboratory heads, sometimes when the head has had nothing to do with the work, or has even been in opposition to it.

Hall's book is written in racy style and is bound to meet with distaste in some quarters. It not only has heroes but also villains, of whom Howard Goodman is cast as the archvillain. Indiscretions from scientists who no doubt now regret having been so frank are all gleefully recorded.

As a guide to how scientists really operate, *Invisible Frontiers* is not in the class of *The Double Helix*. It lacks Watson, the scientist-author, his whimsy and his women — the only women in Hall's book are scientists and, for all he says, none of his scientists seem to have partners or indeed any life outside the laboratory. Nonetheless, this is a book that captures a good deal of the spirit of the times and that should enliven many a two-hour plane flight. □

Peter Newmark is Deputy Editor of Nature.

Courtesy Eli Lilly

Theoretical reality

S. Roy Caplan

Water Movement Through Lipid Bilayers, Pores, and Plasma Membranes: Theory and Reality. Vol. 4, Distinguished Lecture Series of the Society of General Physiologists. By Alan Finkelstein. Wiley: 1987. Pp. 228. \$39.95, £36.65.

AN eminent reviewer of the time recommended that Guggenheim's treatise *Thermodynamics* should be subtitled *Pride and Prejudice*. Without implying a parallel, this remark all too frequently comes to mind while reading Dr Finkelstein's monograph. The pity is that what might have been an excellent source of information is marred by the author's crotchety preconceptions in assessing the contributions of nonequilibrium thermodynamics to his field.

The work is divided into three parts. Part I ("Theory") accounts for virtually half of the text, encompassing both porous and non-porous ("oil") membranes. The remaining half is apportioned evenly between Parts II and III. Part II is devoted to a detailed treatment of the pores generated in lipid bilayer membranes by antibiotics, while Part III deals with the red cell membrane and with epithelia. The latter half of the book contains the most critical assessment of data on physiological water movement available. It is, however, coloured by the views expressed in Part I.

As Dr Finkelstein points out, there is currently no satisfactory theory to describe solvent flow in pores of molecular dimensions. He avoids mentioning that the concept of pressure in such pores is ill-defined, but rather presents us with explanatory diagrams displaying pressure profiles. These intuitive schemes may certainly provide a mental crutch. For the author, however, they constitute reality, prompting him to castigate nonequilibrium thermodynamics for its ability "to obfuscate the physics underlying osmotic transport through a porous membrane". Thus osmotic flow across a membrane separating solvent and a solution of an impermeant solute, both at the same pressure, is said to be driven simply by a pressure gradient within the pore. If this is the whole story, surely the opposing *infinite* pressure gradient (as drawn) at the pore exit also needs explicit consideration.

The picture is extended to the well-known experiment by Meschia and Setnikar supposedly demonstrating water flow in the wrong direction across a collodion membrane separating moderately concentrated urea and dilute dextran at the same pressure. Osmosis proceeds from urea to dextran, against the chemical potential gradient for water. The notion that solute flux may be coupled to, and

hence may drag, solvent flux "completely distorts reality" in Dr Finkelstein's eyes. Not so, he says, water drags urea because the pressure gradient within the pore forces the solution towards the dextran side. How it forces it out of the pore against the large opposing pressure gradient is left for the hapless reader to figure out.

In case of any lingering doubts, Dr Finkelstein asks us to consider a huge sewer pipe with the urea solution at one end, under slight pressure, and pure water at the other. This is claimed to be "exactly comparable" to what occurred in the Meschia and Setnikar experiment. But the comparison is preposterous. The sewer pipe is not of molecular dimensions, and

no opposing pressure gradient exists anywhere. If Dr Finkelstein had read the papers of Kedem and Katchalsky he quotes more closely, he might have noticed that the force driving volume (not solvent) flow in their formulation is the overall pressure difference less the osmotic pressure difference of impermeant solute. This correctly describes the volume flow in both the collodion membrane and the sewer pipe. Is this obfuscation?

Nevertheless, I would urge all interested in this problem to buy Dr Finkelstein's book. Nothing better is around. □

S. Roy Caplan is a Professor in the Department of Membrane Research, The Weizmann Institute of Science, 76 100 Rehovot, Israel.

Fighting fit

John R. Krebs

Animal Conflict. By Felicity Huntingford and Angela Turner. Chapman & Hall: 1987. Pp. 448. Hbk £30, \$59.95; pbk £14.95, \$27.50.

THIS useful introductory book covers much the same ground as Roger N. Johnson's admirable *Aggression in Man and Animals* (W.B. Saunders, 1972), although it is of course much more up to date. It gives a balanced review of aggression and related behaviours in the animal kingdom and considers various theories about both the causes of aggression and its functional significance. The final chapter moves with due caution from animal studies to human aggressive behaviour. Probably the main messages from the book are: (a) in wild animals aggressive behaviour is not a pathology but part of normal adaptive behaviour; (b) different kinds of aggression (predatory, defensive, competitive) may serve different functions and have different causes; and (c) simple general theories on causes of aggression, popular in the past, are likely to be wrong.

In the 1960s, perhaps not unrelated to the peace movement and rise of flower power, there was a fashion for global theories of human aggression: it was argued that aggression is caused by frustration (never thwart your kids and they will grow up as peace-loving citizens), by pain (do away with corporal punishment and kids will stop being aggressive), by copying role models (cut out violent television movies) and so on. We now know that all these ideas are too simple as overall theories of aggression. A widely publicized ethological contribution to 1960s theorizing was the one so forcefully and eloquently expressed by Konrad Lorenz in his classic book *Das sogenannte Böse* (Borotha-Schoeler, 1963), tamely translated into English as *On Aggression* (Methuen, 1966). According to the

Lorenzian view, aggression in animals and man is the expression of an inbuilt drive that wells up inside and has to find release in one form or another. The longer it is since you had a meal, the hungrier you feel, and, according to Lorenz, aggressive feelings build up in the same way since your last fight. Lorenz's view, based in part on the correct notion that aggression, like feeding, is normally an adaptive and essential component of an animal's daily life, is still very influential. Only a couple of months ago I read a newspaper article on football crowd violence which argued that to suppress violence on the terraces would be merely to cause it to erupt elsewhere — the classic Lorenzian theory.

Sidestepping the difficult issue of how and if at all football hooligan aggression is related causally or functionally to other kinds of fighting, one thing is clear: ethological research has shown during the past 20 years that the Lorenzian drive model of aggression in animals is wrong. As Huntingford and Turner make clear, aggression in animals is a form of competition (usually with members of the same species) for limited resources of food, mates or shelter. The evidence shows that the extent to which animals compete by fighting for resources, as opposed to competing by non-aggressive means such as simple exploitation, depends on ecological circumstances. Both within and between species, the extent of aggressive competition depends on the distribution of resources: when they are localized, animals tend to defend them and fight; when resources are scattered or occur unpredictably in time and space, animals tend not to compete by aggression. If anyone wants to use ethological arguments as a starting point for building a new general theory of aggression, this view of aggression as a flexible strategy, rather than the notion of an inevitable internal build-up, should be the starting point. □

John R. Krebs is a Lecturer at the Edward Grey Institute of Field Ornithology, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Try to see things differently

Stuart Sutherland

Vision, Brain, and Cooperative Computation. Edited by Michael A. Arbib and Allen R. Hanson. MIT Press: 1987. Pp. 730. \$65, £58.50.

ONLY 20 years ago it was possible to keep up with all the research published on vision, with the possible exception of colour vision — always an arcane speciality. Today there is so much work done in each subdivision of the field that it is impossible to keep abreast of it all: indeed it has become difficult to cope adequately even with one of the many subdivisions such as motion perception, stereopsis, eye movements, the extraction of lines and edges from the retinal image, the use of motion to form a three-dimensional description of an object, the ability to construct a representation of a scene from glimpses obtained through successive fixations, and the recognition of patterns and objects.

Vision, Brain, and Cooperative Computation deals with all these topics and many more, and as the title implies it does so from the point of view of psychology, neurophysiology and artificial intelligence. The book is not for the faint hearted, for it would take a polyoptacist to understand in depth its highly technical chapters covering such a wide range of phenomena and ways of thought. Nor do the editors, Michael Arbib and Allen Hanson, do much to help, for the graphs are often poorly labelled and captioned, and much of the writing is slovenly. Take for example the sentence "The level of formulation is in terms of symbolic constraints and algorithms for solving them": surely one does not solve constraints though one may use constraints to solve problems. Also, the book is so full of

abbreviations that parts of it are more akin to Stock Exchange listings than to prose. It is sometimes difficult to separate what is new from the frequently occurring but magisterial statements of the banal, such as "I suggest that the cortex has adopted representational and computational strategies that make the computation of invariants efficient", a view with which few are likely to disagree.

Nevertheless — and despite the fact that the conference on which the book is based took place four years ago — most workers on vision will find scattered through its pages nuggets of information with which they are unfamiliar. Sparks and Jay argue that in the superior colliculus, which is involved in the control of eye movements, cells code for the target position of the eye in its orbit, not simply for the retinal distance and direction from the fixation point of the target to be fixated. If the target is briefly illuminated and an eye movement occurs before the fixation movement to the target, account is taken of the new position of the eye.

A more dramatic and better known neurophysiological result is Spinelli's finding that cats exposed only to vertical lines on one eye and only to horizontal on the other develop central cells that are maximally stimulated by a vertical line on the former eye and a horizontal on the latter. This result is important because it demonstrates conclusively that learning can occur at this level of the visual system — the combination of lines is one that cannot occur in nature.

On the psychological side, Weisstein and Wong review some elegant experiments showing that judgements of orientation are much better for lines embedded in a part of the scene that is taken as figure than for lines on what is taken as background. Moreover, while sharp lines are more readily detected when inside a figure, blurred lines are more readily detected when in the background. In both cases the lines were fixated, so these results are not due to differences in retinal

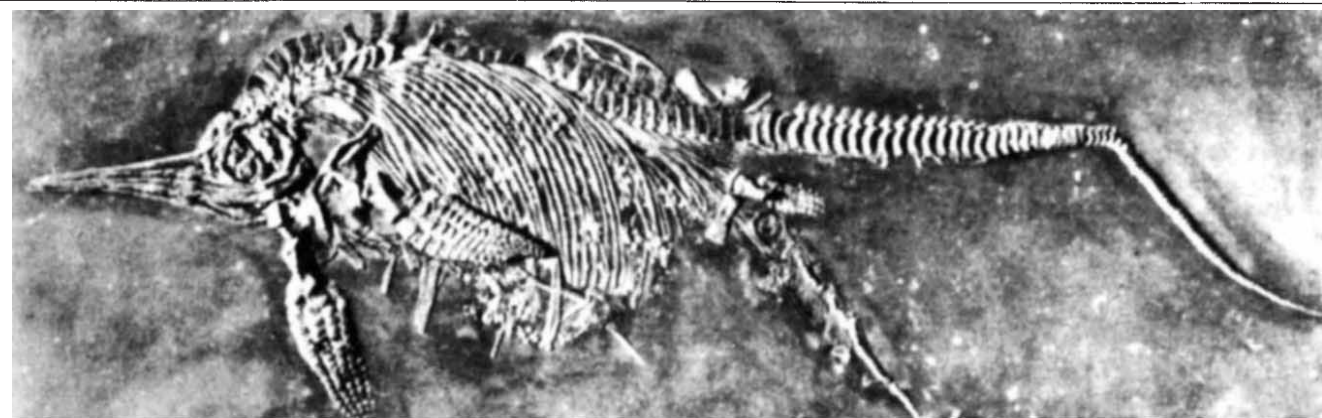
position. Under most circumstances fine detail within a figure is important to us, whereas a general impression of the background will suffice. It is nevertheless remarkable that the visual system has evolved to meet these task requirements in so subtle a way.

Readers' hopes will be aroused by a sentence beginning "In an informal study of the motor skills involved in drinking beer . . .", but alas the findings have no practical application, particularly as they concern only the replacement of the beer mug on the table, not the action of picking it up. In replacing it, there is a fast movement to a point above the table followed by a slower movement that lowers it gently on to the table. The distance above the table at which the fast movement ends increases with increased alcoholic intake, thus allowing a greater margin of error — again a nice example of the way in which the human brain adapts, though in this case the adaptation is presumably due to learning not evolution.

The chapters on artificial intelligence are so opaque that it is hard to know what to make of them. One presents a network for storing the representations of events in their temporal order, but it gives no consideration to the factors that influence people's memory of temporal order. Another develops a suggestion of Alan Cowey's that the reason why different properties of the visual image are coded in different central cells is that far more cells would be needed if there were different cells coding for each possible combination of properties. Rightly or wrongly I felt that many of these contributions were merely restating a problem rather than attempting to solve it.

Like most books of collected papers, this one fails to make a satisfying meal. But it does provide quite a few tid-bits, some of which are more easy to digest than others. □

Stuart Sutherland is Director of the Centre for Research on Perception and Cognition, University of Sussex, Brighton BN1 9QG, UK.



R. Wild. State Museum, Stuttgart

Freeze-frame — a fossil skeleton of a Liassic ichthyosaur in the process of giving birth c. 200 million years ago. The photograph is taken from *Vertebrate Paleontology and Evolution* by Robert L. Carroll, which examines each fossil group in terms of classification, skeletal and functional anatomy, distribution and ecology. The book has over 1,700 illustrations, is just published by W. H. Freeman and costs \$52.95, £47.50.

Rumblings from Hawaii

Floyd W. McCoy

The Volcano Letter. Reprint edition edited by Richard S. Fiske, Tom Simkin and Elizabeth A. Nielsen. *National Museum of Natural History/Smithsonian Institution Press**, 1987. 3kg. \$30.

SEVENTY-FIVE years ago, Thomas A. Jagger settled in Hawaii to study volcanoes; he had become interested in volcanic hazards after seeing the devastation of St Pierre, Martinique, following the 1902 eruption. The Hawaii Volcano Observatory (HVO) was thus in business. To raise funds, Jagger circulated a weekly newsletter to donors — *The Volcano Letter*. The newsletter described tilt measurements and eruptions on Kilauea and Mauna Loa, seismic activity throughout Hawaii and how to make a seismograph; it discussed all aspects of volcanism; it chronicled Jagger's travels; and it catalogued world-wide seismic and volcanic activity. *The Volcano Letter* was thus a curious mix of comment, data and observations.

The Volcano Letter lasted 31 years. It was published in various formats depending upon changing financial circumstances and publishers. Upon Jagger's retirement in 1940, Ruy Finch became editor. With Finch came more scholarly articles, less chatty and with references (an equation printed in 1941, in issue No. 473, signals writing directed to a new audience). Gordon Macdonald followed as editor until 1955 when the last issue was printed. *The Volcano Letter* was left to reside in libraries, increasingly forgotten except by those devoted to Hawaii and its volcanoes.

No more. The entire publication, 530 issues, has been reprinted in one bound volume, fully indexed, to coincide with the seventy-fifth anniversary of HVO. What delightful reading it is — here is a history, a genealogy of research on Hawaiian volcanism. Read Jagger's "Native Superstition" (No. 431) after the bombing of the 1935 Mauna Loa flow heading towards Hilo: the Hawaiians predicted that in "anger the goddess [Pele] would immediately vent her wrath upon [Hilo] and cause widespread destruction. [But] an aged native . . . stopped the flow by direct supplication to the goddess. This gentleman claims direct descent from the goddess herself . . .".

Here are Jagger's predictions of doom for Hilo by either lava flow or tsunami, or both, that culminated in the 1946 tsunami disaster. He pleaded for a Pacific warning

* Orders to E. Nielsen, NHB Mail Stop 119, Smithsonian Institution, Washington, DC 20560, USA.

system, and for lava barriers above Hilo, insisting that "The shadow of the lava menace is just as removable as the shadow of malaria, tuberculosis or yellow fever, the noxious mosquitoes . . ." (No. 443).

The Volcano Letter also contained narratives of eruptions. For instance, the 1880–1881 Mauna Loa flow (No. 520): "We cooked every meal on the lava . . . Occasionally our frying pan floated down with the lava . . . the flow crossed [a] stream, forming a natural . . . bath, from cold to scalding hot". Or this account from Kilauea in 1936 (No. 441): "A thousand people at the active crater . . . are standing . . . on the edge of a vast circular pit . . . motionless . . . a concourse of fire worshippers. All are looking down, over a ring of precipice 800 feet high" — some things never change.

Equally fascinating are Jagger's comments on continental drift; or on the need

to explore the seafloor: "Why go to remote lands when the greatest discoveries in the history of science and invention are within . . . a sampan sail of thirty knots from Hawaii" (this comment was made in 1925, in No. 49).

Wonderful stuff. Here are Jagger, Finch, Powers, Stearns, Wentworth, Stone, Palmer, Macdonald and others talking about their science as much as they are reporting it. But primarily here is a chronicle of one man, Jagger, and his times — a scientist and his interests, an observatory and its development. At \$30, a copy of the book should be in all libraries serving geologists and geophysicists concerned with volcanism or the history of science. □

Floyd W. McCoy is an Adjunct Research Associate Scientist at the Lamont–Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA.

Sensitive science

Martin Smith

Biosensors: Fundamentals and Applications. Edited by Anthony P. Turner, Isao Karube and George S. Wilson. *Oxford University Press*: 1987. Pp.770. £60, \$110.

BIOSENSING is a new field of science which pulls together an enormous range of disciplines, and which is likely to have a considerable effect on clinical testing, environmental monitoring, industrial process control and battlefield early-warning systems. The breadth of the subject and the diversity of the basic science training of existing and prospective practitioners mean that there is the danger of the wheel being re-invented and that constraints on the performance of actual biosensing devices are not appreciated. In addition there is the need to move outside the laboratory and develop sensors that can be manufactured and can live up to the rigours of the proposed operating environment. This is the backdrop to the scene into which *Biosensors: Fundamentals and Applications* makes its entry.

The aim of the editors of this international multi-author text is "to provide the first advanced and comprehensive treatise on the subject of biosensors". In seeking to achieve this aim, the range of subjects covered includes all those that can even loosely be described as biosensing, and there are also chapters on related technology areas such as DNA probes and genetic engineering. All of the contributions are by respected specialists and are well written, and in each area both the fundamental science and the most recent attempts at practical realization are presented. No attempt is made to make

judgements as to which areas of technology are most likely to succeed, which is a realistic approach in view of the embryonic nature of much of the work. However, the level of coverage is such that the skilled reader is well placed to reach his own conclusions.

The book leads naturally from a discussion of principles, through detailed chapters on the basic methods of biosensing (bioelectrochemical, acoustic, calorimetric and photometric), to accounts of instrumentation aspects and commercialization. As would be expected, over half the text is devoted to the many different electrical and electrochemical methods which have been investigated. In many cases the approach taken in expounding each technique is not new, but even the expert reader will find it helpful to have discussion of them brought together in a single tome. Although the coverage is academic and comprehensive, if read selectively the text is nonetheless suitable as a general introduction for the non-specialist reader.

No one can tell where biosensor technology will be in ten years time. The editors of this book make no predictions and neither would I. But it is highly likely that this work will still be a valuable foundation text whichever way the technology progresses. □

Martin Smith is Head of the Sensors Section at Unilever Research, Colworth Laboratory, Sharnbrook, Bedford MK44 1LQ, UK.

• A Discussion on Biosensors, edited by M. Akhtar, C. R. Lowe and I. J. Higgins, is presented in *Phil. Trans. R. Soc. B* 316, 1–181 (1987). This issue is also published as a cloth-bound book entitled *Biosensors*, available to subscribers to *Philosophical Transactions* at £19.80 (UK), £21 (elsewhere), and generally available at £33 (UK), £35 (elsewhere) from the Royal Society, 6 Carlton House Terrace, London SW1Y 5AG, UK.

Vostok ice core: a continuous isotope temperature record over the last climatic cycle (160,000 years)

J. Jouzel*, C. Lorius†, J. R. Petit*†, C. Genthon*, N. I. Barkov‡, V. M. Kotlyakov§ & V. M. Petrov‡

* Laboratoire de Géochimie Isotopique—LODYC (UA, CNRS 1206), CEA/IRDI/DESICP/DPC, 91191 Gif sur Yvette Cedex, France

† Laboratoire de Glaciologie et Géophysique de l'Environnement, BP 96, 38402 Saint Martin d'Heres Cedex, France

‡ Arctic and Antarctic Research Institute, Beringa Street 38, 199226 Leningrad, USSR

§ Institute of Geography, Academy of Science of the USSR, 29 Staronometny, Moscow 109107, USSR

A continuous deuterium profile along the 160,000-year Vostok ice core (Antarctica) is interpreted in terms of atmospheric temperature changes. This climatic record is the awaited terrestrial complement of the deep-sea records supporting the existence of a relation between the Pleistocene climate and orbital forcing.

ON the timescale of the successive glacial-interglacial episodes that have characterized the Earth's climate for about the past million years, quantitative reconstruction of climatic parameters essentially relies on oceanic records, most notably isotopic and faunal studies of deep-sea core sediments. These records provide time series of the continental ice volume and of sea surface temperatures¹⁻³ and have convincingly demonstrated that there is a link between the Pleistocene ice ages and the variations in the elements of the Earth's orbit. By contrast, the climatic conditions that prevailed over continental areas are far less well documented and rarely on a quantitative basis. However, such continental time series are essential to an understanding of how the atmosphere-ocean-cryosphere climatic system is operating on this timescale.

In this context, the 2,083-m ice core⁴ recovered by the Soviet Antarctic Expeditions at Vostok (East Antarctica) is of fundamental importance because it fully covers the last glacial-interglacial cycle, back to the ice age that preceded the last interglacial (160 kyr BP) and has been essentially undisturbed by flow conditions. It allows access to many climatic and climate-related parameters. These potentials are illustrated from the ^{18}O data we have recently published⁴, from ^{10}Be measurements^{5,6}, from the aerosol concentrations⁷ and, in a companion paper, from the CO_2 profile⁸.

Our first isotope dataset was largely discontinuous over the past 100 kyr (only 7% of the core was analysed), and continuous below that. Sampling of the ice was completed later in the field and complementary isotopic measurements were made. We present here continuous data for deuterium throughout the core. We shall focus first on the interpretation of this profile in terms of temperature and then examine this continuous record in the frequency and time domains. We shall compare the ice record with the oceanic record, with a special emphasis on the Last Interglacial Period. For the first time an ice core reveals the large 100-kyr glacial-interglacial cycle and concentration of variance near the Earth's orbit tilt and precession frequencies.

Continuous deuterium profile

The samples were analysed both for their ^{18}O and deuterium contents. Although uncommon in the deep-ice-core literature⁹⁻¹², the present interpretation is based on the deuterium record (denoted δD and expressed in per mil with respect to standard mean ocean water (SMOW)). This is justified because δD is, as discussed later, a slightly better indicator of temperature change than $\delta^{18}\text{O}$.

The basic characteristics of the Vostok site are given in ref. 4. The new samples were taken along 1-m ice increments (instead of 2 m) and were processed as described in ref. 4. The data are now available on an almost continuous basis on the 2,083-m

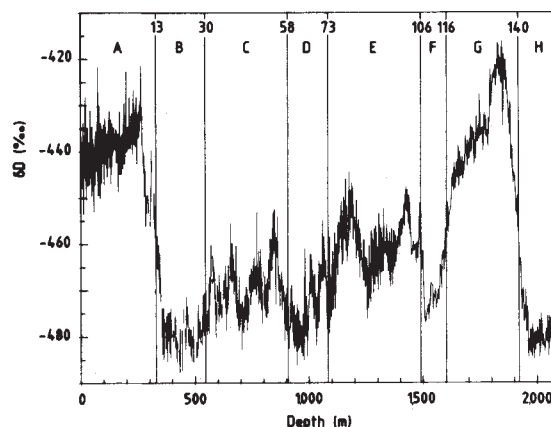


Fig. 1 Deuterium content (in ‰ w.r.t. SMOW) against depth in the Vostok ice core with successive stages as defined in ref. 4 and the ages corresponding to the limits between these stages.

core below 138 m and a complementary dataset was obtained on a new adjacent 279-m core. The total ice recovery is about 85% with only one long (8 m) missing series. Data obtained from the two cores (with an accuracy of 0.3‰ w.r.t. SMOW) are in excellent agreement over their common part (within 1‰ for the average values) and the two series of results were combined to obtain the deuterium profile shown in Fig. 1 as a function of depth.

There are obvious similarities between the present δD and the previous $\delta^{18}\text{O}$ Vostok record with successive warm and cold stages indicated in Fig. 1 and designated by letters from A to H (ref. 4). In the upper part of the figure are also reported the ages corresponding to the limits between stages, following the Lorius *et al.*⁴ chronology. This chronology, established with a two-dimensional ice-flow model and taking into account the change with time of snow accumulation, is discussed in further detail in the last section.

Isotope temperature record

Preliminary estimates of atmospheric temperature changes in the Vostok area were given from the available $\delta^{18}\text{O}$ data⁴. For completing this approach from the continuous δD record, it is useful to examine the validity of this reconstruction from central East Antarctic deep ice cores^{4,11}, which relies on both experimental and theoretical arguments.

The deuterium content distribution in recent snow is well documented over East Antarctica, in particular along the

Dumont d'Urville–Dome C axis¹³. Over a large range of temperature (–20 to –55 °C) there is a linear relation between the annual averages of Θ_s , the surface temperature, and of the snow deuterium content with a δD – Θ_s slope of 6‰ per °C, which approximates very well all the data along this axis. As already noted by Lorius¹⁴, there are some regional features in the isotope distribution over Antarctica, but this value is in the range of that which we recently determined (unpublished results) along the Vostok–Mirny axis (slope of ~7‰ per °C). These slopes are also in the range of that encountered in other parts of Antarctica and in Greenland with an average value of 0.7 ± 0.3 ‰ per °C in $\delta^{18}O$ (ref. 15) (that is ~5.6‰ per °C in δD).

From a simple one-dimensional isotope model, Jouzel and Merlivat¹⁶ have demonstrated that there is agreement between these observations and theoretical predictions, provided one takes as the formation temperature (Θ_i) that prevailing just above the inversion layer¹⁷. There is, on an annual basis, a linear relation between Θ_i and the surface temperature, Θ_s : $d\Theta_i/d\Theta_s = 0.67$. This model accounts for the kinetic fractionation that occurs because snow formation takes place in an environment supersaturated over ice. Indeed, the classical Rayleigh model assuming isotopic equilibrium leads to temperature–isotope gradients that are too high over central East Antarctica. These gradients are reduced when the kinetic effect is considered, down to values around 6‰ per °C for δD (with respect to Θ_s), in good agreement with observations.

This slope reduction is relatively important for $\delta^{18}O$ (up to 50%) and relatively marginal for $\delta D \leq 20$ ‰). Obviously, a change with time in the relative strength of this kinetic effect, possibly associated with changes in saturation conditions, could partly obscure the relation between the isotopic content of polar precipitation and its temperature of formation. Changes in the kinetic effect are isotopic noise in the sense that they are not directly temperature-related. Their relative influence is weaker for δD than for $\delta^{18}O$. The same is true for the kinetic effect associated with sea-surface evaporation, which is dependent on relative humidity^{18,19}. It influences the isotopic content of the vapour leaving the oceanic surface and eventually that of the resulting precipitation, but again relatively less for δD than for $\delta^{18}O$. This is why, as far as both sets of data are available, we prefer to use the δD record for temperature reconstruction.

The distribution of water isotopes in modern precipitation was also recently investigated by two different groups^{20,21} using General Circulation Models of the atmosphere. A simulation of the full annual cycle was performed with the GISS (Goddard Institute for Space Studies) model²¹, and the results can be directly compared with the Antarctic data. The deuterium minimum is well represented over central East Antarctica. The predicted deuterium–surface-temperature gradient is 6.5‰ per °C, very close to the observed one.

Based on these considerations, the past surface-temperature estimate is obtained with a gradient of 6‰ per °C after correcting the Vostok deuterium profile for the change in δD of the oceanic water (the correction is based on the stacked $\delta^{18}O$ record of ref. 3). Figure 2a shows the isotope surface-temperature record. Superimposed on the detailed record obtained after equally spaced resampling every 100 yr is a smoothed curve in which much of the very high frequency oscillations are filtered out. This surface-temperature record is given as a difference with respect to the present value ($\Theta_s = -55.5$ °C). Note that the corresponding atmospheric temperature changes would be approximately two-thirds lower above the inversion layer.

We shall now examine the validity of applying the present deuterium–temperature gradient, well documented and theoretically explained for modern snow on a spatial scale, to an interpretation of the isotopic content of past precipitation, that is, on a temporal basis. This approach requires several conditions to be fulfilled, in particular, that all along the record (1) the influences of factors other than the temperature of formation of precipitation, that is, conditions of temperature and relative

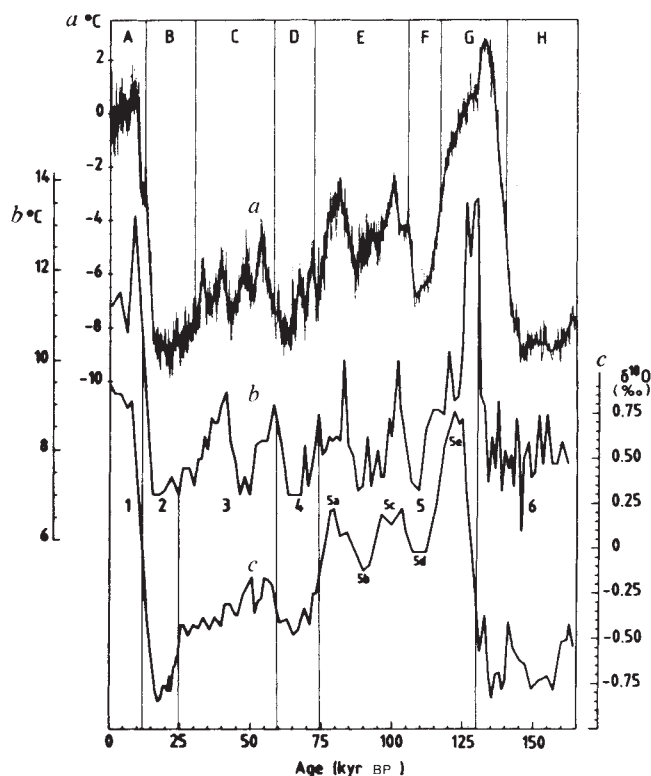


Fig. 2 a, Variation with time of the Vostok isotope temperature record in °C as a difference from the modern surface-temperature value ($\Theta_s = -55.5$ °C). The thin line corresponds to equally spaced values every 100 yr, the thick line to a smoothed curve. Vostok isotopic stages are as in Fig. 1. b, Variation with time of estimated summer sea-surface temperature at the subpolar Indian ocean site RC11-120 (from ref. 3). c, Variation with time of the stacked $\delta^{18}O$ record of ref. 3 (normalized series) with limits between successive marine stages as defined by Martinson *et al.*³, plus an indication of sub-stages 5a to 5e.

humidity in the moisture source regions and shifts of these regions, further dynamical and microphysical history of the air masses, are negligible or at least cancel statistically; (2) the relation between inversion and surface temperatures remains unchanged; (3) there is no additional isotopic noise introduced through variations such as the summer: winter ratio in precipitation amount, which could itself be time-dependent¹⁵.

There are no strong climatological arguments that the above requirements were fulfilled for conditions as different from each other as those prevailing during glacial and interglacial periods. Noting that modern Antarctic and Greenland precipitations are also formed under different climatic conditions, the overall variability of the present surface gradients (~40% from ref. 15, including both spatial and temporal gradients) may be considered an upper limit of the relative error associated with our temperature-change estimates. Indeed, a recent interpretation²² of crystal size variations with depth in terms of surface-temperature change indicates that this uncertainty is probably much smaller. At Vostok the change associated with the last deglaciation is estimated as 11 °C from the crystal size, which compares well with 9 °C in our present interpretation (that is, a relative difference of 20%). Our figure is also consistent with a recent simulation of the Last Glacial Maximum by Broccoli and Manabe²³. Additionally the ^{10}Be profile allowed an estimate of past precipitation⁵. The results are in good agreement with those obtained assuming that precipitation is controlled by saturated vapour pressure⁴; this would not be so if our temperature interpretation of the isotope record was not correct.

Taken together the arguments developed in this section become convincing that much of the isotopic signal has its origin

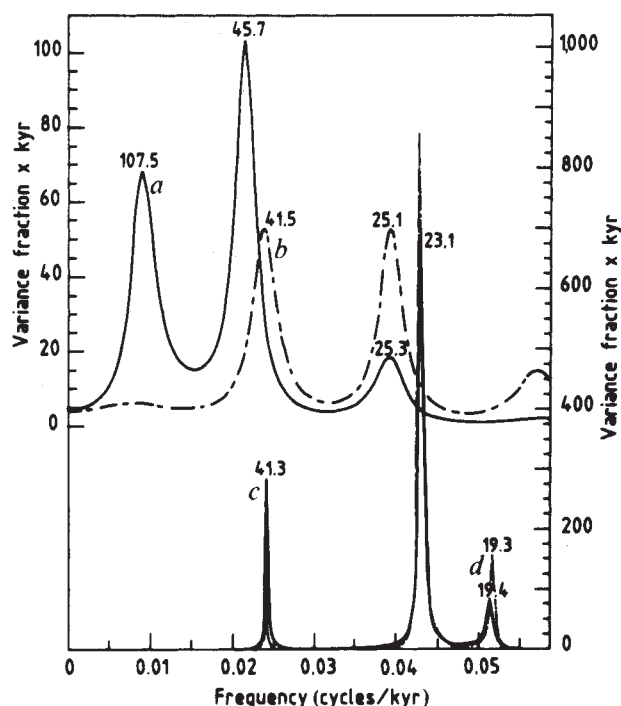


Fig. 3 Variance spectra of the Vostok isotope temperature (*a, b*) and of mid-July insolation at 65° N for the past 160 kyr (*c, d*). The normalized variance density is plotted against frequency in cycles per kyr on linear scales. The left scale is for curves *a* and *b* and the right scale for curves *c* and *d*. Continuous lines refer to the original series (*a, c*) and broken lines to these series after prewhitening by a first-order difference filter (*b, d*). These four spectra were obtained by the maximum entropy method applying the Barrodale and Erickson³³ algorithm with an autoregressive order of 40%.

in the variation of local surface temperature. On the other hand, deviations of up to 20% in the estimate of the temperature change cannot be excluded. Finally, to get the climatically induced temperature change, corrections have to be applied for the changes of elevation either due to the motion of the ice from its origin or to general changes in the thickness of the ice sheet²⁴. We now have new isotopic results from snow surface samples both at Vostok and Ridge B, that is, over the region from which the Vostok ice probably originated over the last climatic cycle⁴. The two average deuterium values for about the last century fall within a few per mil (−438‰ at Vostok and −435‰ at Ridge B). This confirms that there is no need for a correction linked with the origin of the ice. The available data on ice cap thickness (D. Raynaud, personal communication) show that there are variations in the total gas content of the ice, a parameter directly related to the elevation of the ice cap, which are roughly in phase with the deuterium changes. Taking this effect into account would amplify the amplitude of our estimated temperature variations but not the shape of the profile. This aspect is not taken into consideration in the present interpretation but will need to be kept in mind in a final assessment of the data.

Spectral analysis of temperature record

The Vostok ice offers a unique opportunity to examine from a continental record the link existing between a time series for atmospheric temperature and the astronomical forcing that drives the amount of insolation received at any latitude and time. Among the orbital parameters, the obliquity of the Earth's axis (period of ~41 kyr) and the precession of the equinoxes (periods at 23 and 19 kyr) are the most important as they redistribute the available energy between latitudes and months^{25–28}. For example, there is about a 20% difference in the mid-July insolation received at 65° N (a key latitude in the

classical theory of the ice ages) between a time of high or low seasonal contrast²⁵. On the other hand, the total insolation received by the planet, only influenced by the Earth's eccentricity (periods at 100 and 400 kyr) varies by only about 0.3% over the past 160 kyr. The ~100-kyr oscillation that dominates the climate record over the past million years is generally assumed to result from a nonlinear response of the ice sheets to orbital forcing rather than directly from eccentricity changes^{28–31}.

Here we focus the spectral analysis on the obliquity and precessional frequency bands. The higher frequency spectra, which also contain information about the behaviour of the climatic system^{30,31}, will be examined elsewhere. For this analysis we followed the strategy described by Berger and Pestiaux²⁷, who suggested the use of a combination of different methods, each with its own advantages and drawbacks.

The commonly used Blackman–Tukey³² technique has well-defined statistical properties and is reliable for spectral-amplitude information but has poor frequency resolution. It is not very appropriate in our case of a record that is short with respect to the periods of interest. Even using a large number of lags (50%), there is no clear indication of spectral peaks.

We improved the frequency resolution using the Maximum Entropy (ME) method by applying a least-squares algorithm³³ appropriate for short records. Figure 3 (curve *a*) shows the Vostok spectrum. The spectrum of mid-July insolation at 65° N for the past 160 kyr (curve *c*) illustrates the quality of this method, which clearly separates the 23-kyr and 19-kyr precessional peaks. To depict the higher frequency part of the record more clearly, the Vostok series was prewhitened with a first-order difference filter, thus removing variance at low frequency (curve *b*). Curve *d* shows the corresponding spectrum for the 65° N insolation.

These spectra were calculated with an autoregressive order of 40%. We checked that the position of the peaks is fairly stable (within ~±3%) over a large range of values of this parameter (from 35 to 50%). The unprewhitened Vostok spectrum (Fig. 3, curve *a*) shows well-defined peaks at 107.5, 45.7 and 25.3 kyr. A much larger part of the variance is concentrated in the ~40-kyr peak than in the 20-kyr peak. After filtering of the series (curve *b*) the 45.6-kyr peak is shifted towards higher frequencies (41.5 kyr), probably because its position was influenced by the strong 100-kyr component. In contrast, the position of the 41-kyr peak for 65° N insolation is not affected by the prewhitening (curve *d*) because there is no such strong 100-kyr component.

The Vostok peaks at ~40 and ~20 kyr are on the lower frequency side of the 41-kyr obliquity component and of the 23-kyr precessional main component but the differences are well within the dating accuracy, which is in the range 5–10% (ref. 4). Note here that consequence of claiming that δD is a better temperature indicator than is $\delta^{18}O$ should be to correct the timescale. The influence of using δD rather than $\delta^{18}O$ in the age model of ref. 1 never exceeds 2%. This is well below the dating uncertainty as the influence of ice thickness probably changes.

Within these limits, our spectral analysis demonstrates convincingly that, apart from the ~100 kyr signal, the Vostok temperature record is dominated by a strong component near 40 kyr previously seen from visual inspection of the record⁴. It also suggests that it is influenced to a lesser degree by a component slightly above 20 kyr. The two frequency bands may be associated with the obliquity and precession cycles respectively. This result is quite remarkable if we recall that the Vostok dating is based only on a glaciological model⁴. When deep-sea core profiles of $\delta^{18}O$ are analysed without any assumption of a relation between an insolation curve and the isotopic record, they generally do not show better defined peaks in the obliquity and precessional bands. For example, Berger and Pestiaux²⁷, from the study of 10 cores dated in this way, showed that the variability in the position of the 41-kyr and 23-kyr peaks is

≥ 5 kyr. At the least, the Vostok atmospheric temperature record appears to be the awaited terrestrial complement of the deep-sea cores to support the role of astronomical forcing in determining the last Pleistocene climate.

Comparison with marine climatic record

Figure 2 (curves *c* and *b*) shows time series of the estimated continental ice volume change and of a Southern Hemisphere sea-surface temperature record for the past 160 kyr. They represent respectively the stacked SPECMAP $\delta^{18}\text{O}$ record of Martinson *et al.*³, with the 'standard' stratigraphy for deep-sea sediments and the estimated summer sea-surface temperature from the very well documented^{1,3} and geographically relevant subpolar Indian Ocean core RC11-120. For these two series the chronology established by Martinson *et al.*³ stems explicitly from orbital tuning. We discuss here the salient characteristics of the Vostok record and its comparison with these two marine climatic series, going back in time from the Holocene to the last Interglacial.

The warmest part of the Holocene occurred just at the end of the last deglaciation, which is also seen in RC11-120 (temperature maximum occurring at $\sim 9,000$ kyr BP) and was previously observed in the Dome C core¹¹. At Vostok the surface-temperature shift between the Last Glacial Maximum (LGM) and average Holocene conditions is about 9°C . This transition is a two-step process with two warming intervals interrupted by a slightly colder one lasting ~ 1 kyr (from ~ 12 to 11 kyr BP in our chronology). The temperature amplitude of this cold episode is $\sim 2^\circ\text{C}$. There is some missing ice in this part of the record, but there is still no doubt that a cold oscillation exists in this Vostok transition as in Dome C¹¹ and possibly Byrd⁹. The relationship between this Antarctic event and the well-documented European Younger Dryas (see ref. 34 for an extensive review) well recorded in the Greenland deep ice cores¹² is examined in Jouzel *et al.*³⁵. According to the best dating so far there is no significant lead or lag between the three climatic curves of Fig. 2 at the beginning of this transition from the LGM towards the Holocene, which takes place in the three records at ~ 15 kyr BP.

We had interpreted⁴ stages B, D and F (which correspond to marine stages 2, 4 and 5d) from $\delta^{18}\text{O}$ as being very similar. Figure 2 (curve *a*) shows a decreasing trend of the temperature minima from F to B, with stage F being $\sim 2^\circ\text{C}$ warmer than the full glacial conditions (stage B). This difference in our interpretation of stage F partly relates to the fact that the deuterium excess ($d = \delta D - 8\delta^{18}\text{O}$), largely depending upon the above-mentioned kinetic effects, is then particularly high compared with that recorded during the other cold stages (the continuous deuterium-excess profile will be discussed elsewhere). In fact, conditions equivalent to the LGM were encountered only at the end of the penultimate glacial (stage H).

The two C and E interstadials are now thoroughly documented, demonstrating in each case a well-defined internal structure.

During stage C there are two double-peak maxima separated by a period of colder conditions (but about 1.5°C above the LGM) at ~ 42 kyr BP. The warmest conditions prevailed at the beginning of this interstadial (~ 53 kyr BP) with temperature up to 4°C above full glacial conditions. There is a reasonable correspondence between these Vostok stage C temperature maxima and the corresponding parts of the oceanic records (marine stage 3). There is in the detailed curve nothing comparable with what is recorded in the Dye 3 Greenland ice core¹², which shows at the timescale of a few decades abrupt and vigorous isotopic changes with an amplitude almost equivalent to that accompanying the Pleistocene-Holocene shift. The lack of such large oscillations in our core is far from proven, given the relatively coarse sampling we used.

Vostok interstadial E is characterized by two well-marked temperature maxima (~ 81 and 100 kyr BP) and some minor

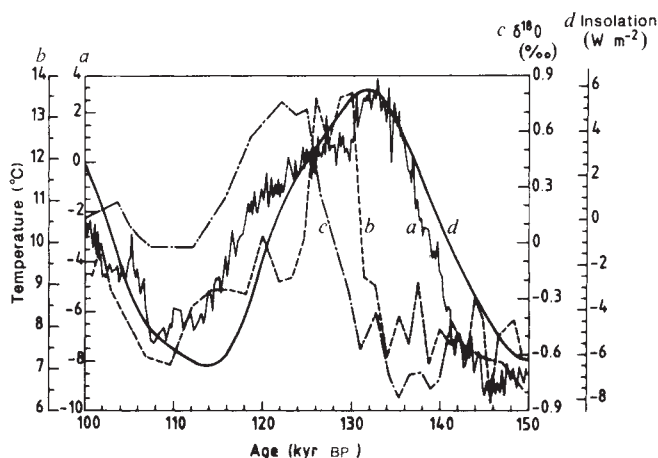


Fig. 4 Variation with time of the three climatic series of Fig. 2 for the period between 100 and 150 kyr BP. *a*, Vostok isotope temperature (thin continuous line); *b*, RC11-120 summer sea-surface temperature from ref. 3 (broken line); *c*, stacked $\delta^{18}\text{O}$ record from ref. 3 (dot-dashed line); *d*, annual insolation at 78°S (thick line). The four curves were scaled, for the sake of comparison, to have about the same amplitude.

additional oscillations. For both events, the temperature culminates at a level $\sim 6^\circ\text{C}$ above the LGM temperature. There is quite clear correspondence between marine stages 5a and 5c, seen both in the $\delta^{18}\text{O}$ and temperature oceanic record, and these two relatively warm periods. Correspondingly, stage 5b agrees well with a period $\sim 2^\circ\text{C}$ colder recorded in the Vostok ice from ~ 95 to 85 kyr BP.

We note here the remarkable correlation ($r^2 = 0.87$) between the marine $\delta^{18}\text{O}$ and the Vostok record down to ~ 110 kyr BP. As the marine $\delta^{18}\text{O}$ record is thought to represent essentially continental ice-volume changes, that is, a parameter of global significance, this high correlation suggests that the Vostok temperature record is, at least qualitatively, of relatively large geographical significance.

Such a general agreement for timescales between the ocean and ice climatic records does not hold for the earlier period of the record (before 110 kyr BP). We shall now concentrate on this period after pointing out that the δD interpretation confirms⁴ that the Last Interglacial contains an interval of ~ 5 kyr warmer ($\sim 2^\circ\text{C}$) than present conditions. For the sake of this discussion parts of the three palaeorecords from 150 to 100 kyr BP are shown in Fig. 4 (curves *a* to *c*). The major difference is in the duration of the Last Interglacial, namely 11 kyr (127–116 kyr BP) for the $\delta^{18}\text{O}$ deep-sea-core stage 5e and about 22 kyr (139–117 kyr BP) for the Vostok stage G. This difference inevitably induces discrepancies between the ice and deep-sea-core chronologies. Some mechanisms like the Antarctic floating-shelves hypothesis³⁶ could explain such a difference. We shall argue instead that these major climatic changes were nearly synchronous on a global scale. First, we assume^{8,37} that apart from astronomical forcing, CO_2 changes that are global have played a climatic role. Second, there is little doubt that the last deglaciation was nearly in phase in both hemispheres³⁸, as discussed earlier in this section, and it is difficult to justify that a different glacial mechanism with a large phase lag had operated during the penultimate deglaciation. Third, as firmly pointed out by Broecker³⁸ and J. Imbrie (personal communication) temperature changes over Antarctica must be nearly synchronous with changes in surface temperature conditions around the Antarctic continent.

The deep-sea-core chronology³ is independently constrained and very well confirmed by dating, in particular, of Barbados high-sea-level terraces, which sets Barbados III level as corresponding to marine stage 5e at 125 kyr BP (refs 39, 40). In disagreement with Barbados terraces chronology, there is an

abundant literature⁴¹⁻⁴³ on Huon Peninsula (New Guinea) sea levels, showing a doubled-peak long interglacial (~25 kyr at mid-peak) with two high sea-level stands at 132 and 120 kyr BP that would fit with the Vostok chronology. Let us follow the CLIMAP project members in their conclusion that these terrace ages must be in error⁴⁴. However, there are apparently reliably dated coral reefs of Haiti that also give a high sea-level stand at 130 kyr BP (ref. 45). Such a possibility is at least not contradicted (the earlier date is 129.9 kyr BP) by recent and more accurate datings using mass spectrometry⁴⁶. These dates leave open the possibility of an earlier interglacial as suggested by the Vostok record. Indeed, as Southern Hemisphere temperature changes are expected to lead sea-level changes by a few kyr (ref. 44) a maximum sea level at 130 kyr BP would be in agreement with the Vostok profile, which shows its temperature maximum at ~133 kyr BP.

In the Vostok chronology underestimation of both the accumulation rate and the thinning function could have led us to overestimate the duration of Vostok stage G (ref. 4). Previously published⁵ and new detailed⁶ ¹⁰Be measurements are consistent with quite similar snow accumulation rates at the peak of the Last Interglacial and during the Holocene. Aerosol input of either continental or marine origin, as estimated from sodium and aluminium measurements, is not very different between the Holocene and the Last Interglacial Maximum⁷, indirectly strengthening this argument although changes in atmospheric circulation or in production rates (for ¹⁰Be) may certainly not be excluded. The thinning function depends on flow parameters, which may be affected by various factors such as concentration of dust and other impurities and crystal size and fabrics⁴⁷. Preliminary experiments suggest that the Vostok interglacial ice may be only slightly more viscous than glacial ice⁴⁸; a corresponding adjustment leads to only small changes in our chronology.

Over this period, relevant Southern Hemisphere insolation curves show a maximum before 130 kyr BP. For example, this is true for November insolation at 60° S, which partly controls the summer extent of sea-ice around Antarctica⁴⁹ and for the annual average of the local insolation shown on Fig. 4 (curve *d*). Indeed, there is a remarkable correlation between this insolation curve and the Vostok temperature over the 150–100 kyr interval. If the Antarctic climate were governed in some way by Southern Hemisphere insolation changes (ref. 50; G. Kukla, personal communication), this would tend to support the Vostok chronology.

Thus we think that there is no convincing argument to modify the Vostok timescale of ref. 4, which is adopted here and in the companion papers^{8,37}. This choice does not mean that we deny the role of orbital variations in influencing the Earth's climate; indeed the application of orbital tuning techniques to the Vostok record is being explored in collaboration with J. Imbrie. But we do claim that the Vostok ice core offers a unique source of information which will be most valuable if dated independently of the deep-sea core record.

Conclusions

We now have obtained a continuous deuterium record along the Vostok ice core that we have interpreted in terms of local surface temperature changes over the past 160 kyr. This is dominated by the large ~100 kyr glacial-interglacial signal with a total temperature amplitude of ~11 °C; the Last Glacial Period is very well documented and it is confirmed that the warmest part of the Last Interglacial Period was about 2 °C warmer than the Holocene. Apart from arguments based on modern surface-snow data and on a dynamically simple isotopic model, such an interpretation is consistent with independent approaches such as that based on crystal-size variations and climate modelling.

A special emphasis is given to the comparison between the Vostok and the marine records. The well-marked similarities

between, in particular, the estimated ice volume and the Vostok temperature suggests that our record is, at least qualitatively, of relatively large geographical significance. Such agreement is not observed before 110 kyr BP owing to discrepancies between the Vostok and marine chronologies. The duration of the Last Interglacial Period in the Vostok record is about twice as long as expected, if the late Pleistocene Earth's climate was governed only by Northern Hemisphere high-latitude summer insolation. Our present strategy is to keep this Lorius *et al.*⁴ timescale, as long as there is no inescapable argument to modify it.

It is remarkable that without any radiometric dates or orbital tuning the Vostok temperature record shows concentration of variance near the obliquity and precession frequency bands, which strongly supports the climatic role of orbital forcing. This aspect and the influence of other forcings (in particular atmospheric CO₂) is examined in the two companion papers^{8,37}. These two studies illustrate ways in which obtaining the first continuous long record of the atmospheric temperature, which is the main achievement of the present study, contributes to a better understanding of glacial-interglacial climatic changes.

We thank all Soviet and French participants in drilling, field work and ice sampling. We acknowledge the Soviet Antarctic Expeditions and the US Division of Polar Programs for the logistic support. We thank R. Chiron and G. Mondet for isotopic determinations. The comparison between the Vostok and marine records has largely benefited from very constructive criticisms of W. S. Broecker, J. C. Duplessy, R. Fairbanks, J. Hays, J. Imbrie, G. Kukla, L. D. Labeyrie, A. McIntyre, R. K. Matthews, J. Morley, and J. White. We have had useful discussions with A. Berger, M. Ghil, S. Joussaume, L. Merlivat, D. Peteet, H. Le Treut, G. Raisbeck and F. Yiou. We also thank J. P. Benoist, E. Cook and J. Portes for their help in data analysis, and J. M. Barnola and D. Raynaud for very useful discussions and suggestions about the presentation of this work. We very much appreciated the review of H. Oeschger. This work was supported in France by the EPF (Expéditions Polaires Françaises), PNEDC (Programme National d'Études de la Dynamique du Climat) and TAAF (Terres Australes et Antarctiques Françaises) and in the USSR and Vostok by Soviet Antarctic Expeditions.

Received 26 March; accepted 17 July 1987.

- Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
- Imbrie, J. *et al.* in *Milankovitch and Climate* Vol. 1 (eds Berger, A. L. *et al.*) 269–305 (Reidel, Dordrecht, 1984).
- Martinson, D. G. *et al.* *Quat. Res.* **27**, 1–29 (1987).
- Lorius, C. *et al.* *Nature* **316**, 591–596 (1985).
- Yiou, F., Raisbeck, G. M., Bourles, D., Lorius, C. & Barkov, N. I. *Nature* **316**, 616–617 (1985).
- Raisbeck, G. M. *et al.* *Nature* **326**, 273–277 (1987).
- De Angelis, M., Barkov, N. I. & Petrov, V. N. *Nature* **325**, 318–321 (1987).
- Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
- Johnsen, S. J., Dansgaard, W., Clausen, H. B. & Langway, C. C. *Nature* **235**, 429–434 (1972).
- Barkov, N. I., Korotkevich, E. S., Gordinenko, F. G. & Kotlyakov, V. M. *I.A.H.S. publ.* **118**, 312–321 (1977).
- Lorius, C., Merlivat, L., Jouzel, J. & Pourchet, M. *Nature* **280**, 644–648 (1979).
- Dansgaard, W. *et al.* *Geophys. Monogr. (M. Ewing Symp. 5)* **29**, 288–298 (1984).
- Lorius, C. & Merlivat, L. *I.A.H.S. Publ.* **118**, 127–137 (1977).
- Lorius, C. in *The Climatic Record of the Polar Ice Sheets* (ed. Robin, G. de Q.) 52–56 (Cambridge University Press, 1983).
- Robin, G. de Q. in *The Climatic Record of the Polar Ice Sheets* (ed. Robin, G. de Q.) 180–195 (Cambridge University Press, 1983).
- Jouzel, J. & Merlivat, L. *J. geophys. Res.* **89**, 11749–11758 (1984).
- Robin, G. de Q. *Phil. Trans. R. Soc. B280*, 143–148 (1977).
- Merlivat, L. & Jouzel, J. *J. geophys. Res.* **84**, 5029–5033 (1979).
- Jouzel, J., Merlivat, L. & Lorius, C. *Nature* **289**, 688–691 (1982).
- Joussaume, S., Jouzel, J. & Sadourny, R. *Nature* **311**, 24–29 (1984).
- Jouzel, J. *et al.* *J. geophys. Res.* (submitted).
- Petit, J. R., Duval, P. & Lorius, C. *Nature* **326**, 62–64 (1987).
- Broccoli, A. J. & Manabe, S. *Climate Dynam.* **1**, 87–99 (1987).
- Raynaud, D. & Lorius, C. *I.A.H.S. Publ.* **118**, 326–344 (1977).
- Berger, A. L. *J. Atmos. Sci.* **3**, 2362–2367 (1978).
- Berger, A. L. & Pestiaux, P. in *Milankovitch and Climate* Vol. 1 (eds Berger, A. L. *et al.*) 83–112 (Reidel, Dordrecht, 1984).
- Berger, A. L. & Pestiaux, P. *Current Issues in Climate Research*, 76–96 (Reidel, Dordrecht, 1986).
- Imbrie, J. & Imbrie, J. Z. *Science* **207**, 243–253 (1980).
- Budd, W. F. & Smith, I. N. *I.A.H.S. Publ.* **118**, 369–409 (1981).
- Le Treut, H. & Ghil, M. *J. geophys. Res.* **88**, 5167–5190 (1983).
- Ghil, M. in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics* (eds Ghil, M., Benzi, R. & Parisi, G.) 347–402 (North-Holland, Amsterdam, 1985).

32. Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Application* (Holden-Day, San Francisco, 1968).
33. Barrodale, I. & Eriksson, R. E. *Geophysics* **45**, 420-432 (1980).
34. Rind, D., Peteet, D., Broecker, W., McIntyre, A. & Ruddiman, W. *Climate Dynam.* **1**, 3-33 (1986).
35. Jouzel, J., Lorius, C., Merlivat, L. & Petit, J. R. *Symp. on Abrupt Climatic Changes* (in the press).
36. Johnson, R. G. & Andrews, J. T. *Paleogeogr., Paleoclimatol. Paleocool.* **53**, 107-138 (1986).
37. Genthon, C. *et al. Nature* **329**, 414-418 (1987).
38. Broecker, W. S. in *Milankovitch and Climate* Vol. 2 (eds Berger, A. L. *et al.*) 687-698 (Reidel, Dordrecht, 1984).
39. Mesolella, K. J., Matthews, R. K., Broecker, W. S. & Thurber, D. L. *J. Geol.* **77**, 250-274 (1969).
40. Fairbanks, R. G. & Matthews, R. K. *Quat. Res.* **10**, 181-197 (1978).
41. Chappell, J. *Bull. geol. Soc. Am.* **85**, 553-570 (1974).
42. Chappell, J. *Search* **3-4**, 99-101 (1983).
43. Aharon, P. *Nature* **30**, 720-723 (1983).
44. Climap project members *Quat. Res.* **21**, 123-224 (1984).
45. Dodge, R. E., Fairbanks, R. G., Benninger, L. K. & Maurasse, L. *Science* **219**, 1423-1425 (1983).
46. Edwards, R. L., Chen, J. H. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **81**, 175-192 (1986).
47. Reeh, N. *Nature* **317**, 797-799 (1985).
48. Pimienta, P., Duval, P. & Lipenkov, V. Ya. *Ann. Glaciol.* (submitted).
49. Kukla, G. in *Climatic Change* (ed. Gribbin, J.) 114-129 (Cambridge University Press, 1978).
50. Bernard, E. A. *Dakar Symposium, Union Internationale pour l'Étude du Quaternaire* (1986).

Vostok ice core provides 160,000-year record of atmospheric CO₂

J. M. Barnola*, D. Raynaud*, Y. S. Korotkevich† & C. Lorius*

* Laboratoire de Glaciologie et de Géophysique de l'Environnement, BP 96, 38402 Saint Martin d'Hères Cedex, France

† Arctic and Antarctic Research Institute, Beringa Street 38, Leningrad 199226, USSR

Direct evidence of past atmospheric CO₂ changes has been extended to the past 160,000 years from the Vostok ice core. These changes are most notably an inherent phenomenon of change between glacial and interglacial periods. Besides this major 100,000-year cycle, the CO₂ record seems to exhibit a cyclic change with a period of some 21,000 years.

ALTHOUGH the first direct CO₂ measurements in the atmosphere were made in the second half of the nineteenth century, atmospheric CO₂ variations have been monitored in a systematic and reliable manner only since 1958. Fortunately, nature has been taking continuous samples of the atmosphere at the surface of the ice sheets throughout the ages. This natural sampling process takes place when snow is transformed into ice by sintering at the surface of the melt-free zones of the ice sheets, with air becoming isolated from the surrounding atmosphere in the pores of the newly formed ice. After pore closure the gas remains stored in the ice moving within the ice sheet. During this natural air sampling and storage process, different mechanisms could alter the original atmospheric composition^{1,2}. But by choosing appropriate sampling sites, ice cores (for example see ref. 1) and experimental methods³, past CO₂ changes in the atmosphere can be determined with high confidence by analysing the air enclosed in the pores of the ice.

Previous results from ice-core analysis have already provided important reliable information on the 'pre-industrial' CO₂ level and the recent CO₂ increase induced by anthropogenic activities⁴⁻⁶. Striking CO₂ changes have also been detected in this way in the ice record covering the last 30-40 kyr⁷⁻¹⁰, including the large CO₂ increase associated with the climatic shift from the Last Glacial Maximum (~18 kyr BP) to the Holocene.

Because of the extremely low temperatures at Vostok (present-day mean annual temperature is -55.5 °C) and the good core quality, the 2,083-m-long ice core recovered by the Soviet Antarctic Expeditions at Vostok (East Antarctica) provides a unique opportunity to extend the ice record of atmospheric CO₂ over the last glacial-interglacial cycle back to the penultimate ice age about 160 kyr ago¹¹. Over this timescale a high correlation is found between CO₂ concentrations and Antarctic climate, with significant oscillatory behaviour of CO₂ between high levels during interglacial and low levels during glacial periods. The CO₂ record also seems to exhibit a cyclic change with a period of ~21 kyr, that is, around the orbital period corresponding to the precession.

Experimental procedure

Gas extraction and measurements were performed with the 'Grenoble analytical setup' described by Barnola *et al.*¹². The

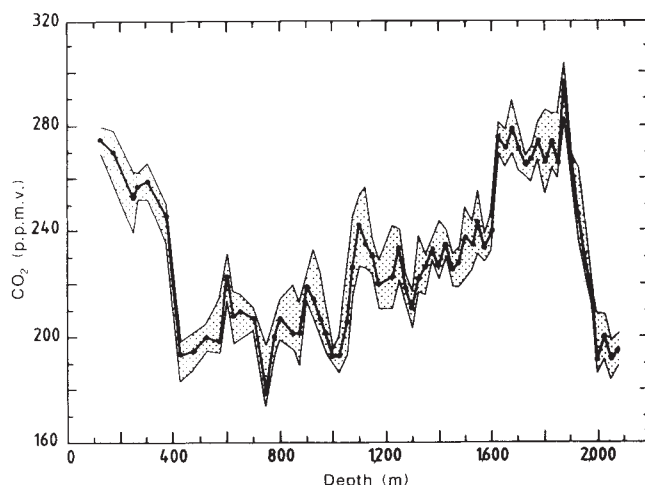


Fig. 1 CO₂ concentrations (p.p.m.v.) plotted against depth in the Vostok ice core. The 'best estimates' of the CO₂ concentrations are indicated by dots and the envelope shown has been plotted taking into account the different uncertainty sources.

method is based on crushing the ice under vacuum without melting, expanding the gas released during the crushing in a pre-evacuated sampling loop, and analysing the CO₂ concentrations by gas chromatography.

The analytical system, except for the stainless steel container in which the ice is crushed, is calibrated for each ice sample measurement with a standard mixture of CO₂ in nitrogen and oxygen. The corresponding accuracy (2σ) is evaluated from the standard deviation of the residuals corresponding to the calibration regression and ranges from 3 to 12 parts per million by volume (p.p.m.v.) for the measurements presented in this article.

We recently discovered an additional error due to our experimental system when a significant amount of water vapour is detected by the gas chromatograph. In such a case, selective CO₂ transport by water vapour (similar to that observed by Neftel *et al.*⁴) back to the ice crushing container is suspected of depleting the CO₂ from the air extracted from the ice and injected

Table 1 Mean CO₂ concentrations in Vostok ice core (best estimates and corresponding uncertainties, against depth and age)

Depth (m)	Age of the ice (yr BP)	Mean age of the air (yr BP)	CO ₂ concentration (best estimate) (p.p.m.v.)	Depth (m)	Age of the ice (yr BP)	Mean age of the air (yr BP)	CO ₂ concentration (best estimate) (p.p.m.v.)
126.4 (Y)*	4,050	1,700	274.5 ± ⁵ / ₁₃	1,274.2 (Y, N)	87,980	84,700	218.8 ± ^{5.5} / _{5.5}
173.1 (N)	5,970	3,530	270.0 ± ¹³ / ₁₃	1,299.3 (N)	89,940	86,680	210.0 ± ⁷ / ₁₀
250.3 (N, N)	9,320	6,800	252.0 ± ¹³ / _{10.5}	1,322.5 (A)	91,760	88,520	221.5 ± ⁴ / ₁₀
266.0 (N)	10,040	7,500	257.0 ± ⁵ / ₇	1,349.0 (N)	93,860	90,630	226.0 ± ⁴ / ₁₀
302.6 (N)	11,870	9,140	259.0 ± ⁵ / ₇	1,374.8 (Y)	95,910	92,700	234.0 ± ⁴ / ₁₀
375.6 (Y)	16,350	12,930	245.0 ± ¹⁰ / ₁₀	1,402.5 (A)	98,130	94,940	226.5 ± ³ / ₁₀
426.4 (N)	20,330	16,250	193.0 ± ⁵ / ₁₀	1,425.5 (Y)	100,000	96,810	236.0 ± ⁵ / ₁₀
474.2 (N)	24,280	20,090	194.5 ± ⁵ / ₁₀	1,451.5 (N)	102,210	98,950	225.0 ± ⁵ / ₁₀
525.1 (Y)	28,530	24,390	200.0 ± ⁵ / ₁₀	1,476.1 (N)	104,410	101,040	229.0 ± ¹⁰ / ₁₀
576.0 (A)	32,680	29,720	198.0 ± ¹⁷ / ₁₇	1,499.6 (Y)	106,610	103,130	238.5 ± ¹⁶ / ₁₆
602.3 (Y)	34,770	30,910	223.0 ± ⁹ / ₁₀	1,526.3 (N)	109,240	105,620	234.5 ± ⁹ / ₁₀
625.6 (N)	36,600	32,800	207.0 ± ¹⁰ / ₁₀	1,547.0 (Y)	111,250	107,650	244.0 ± ¹² / ₁₂
651.6 (Y)	38,600	34,870	210.0 ± ⁵ / ₁₀	1,575.2 (N)	113,850	110,510	233.5 ± ⁵ / ₁₀
700.3 (N)	42,320	38,660	207.0 ± ⁵ / ₁₀	1,598.0 (N)	115,850	112,700	240.0 ± ¹⁰ / ₁₀
748.3 (A)	45,970	42,310	178.5 ± ⁵ / ₁₀	1,626.5 (Y)	118,220	115,290	276.0 ± ⁶ / ₁₀
775.2 (Y)	48,000	44,350	200.0 ± ¹⁸ / ₁₈	1,651.0 (Y, N)	120,170	117,410	271.5 ± ⁹ / ₁₀
800.0 (N, Y, Y)	49,850	46,220	207.7 ± ^{8.4} / ₁₀	1,676.4 (N)	122,100	119,500	280.0 ± ¹⁰ / ₁₀
852.5 (A)	53,770	50,150	201.0 ± ¹² / ₁₂	1,700.9 (N)	123,900	121,430	271.0 ± ⁸ / ₁₀
874.3 (N)	55,450	51,770	201.0 ± ¹² / ₁₂	1,726.8 (N, N)	125,730	123,380	265.3 ± ⁸ / ₁₀
902.2 (N)	57,660	53,860	219.5 ± ¹⁰ / ₁₀	1,747.3 (Y, N)	127,150	124,880	267.0 ± ⁴ / ₁₀
926.8 (A)	59,670	55,780	214.5 ± ¹⁹ / ₁₉	1,774.1 (N)	129,020	126,770	275.0 ± ⁷ / ₁₀
951.9 (A)	61,790	57,800	206.5 ± ¹⁸ / ₁₈	1,802.4 (A)	131,030	128,780	266.5 ± ¹² / ₁₂
975.7 (Y)	63,880	59,770	201.0 ± ⁸ / ₁₀	1,825.7 (Y, N)	132,700	130,460	275.0 ± ¹⁰ / ₁₀
1,002.5 (N)	66,230	62,080	192.0 ± ³ / ₁₀	1,850.5 (A)	134,510	132,280	266.0 ± ⁶ / ₁₀
1,023.5 (N)	68,040	63,960	193.0 ± ² / ₁₀	1,875.9 (Y)	136,450	134,170	296.5 ± ¹³ / ₁₃
1,052.4 (A)	70,470	66,540	205.5 ± ¹² / ₁₂	1,902.0 (Y)	138,660	136,170	266.0 ± ⁸ / ₁₀
1,074.8 (A)	72,330	68,490	226.5 ± ¹¹ / ₁₁	1,928.0 (A)	141,170	138,410	246.5 ± ¹⁰ / ₁₀
1,101.4 (Y)	74,500	70,770	243.0 ± ¹⁸ / ₁₈	1,948.7 (A)	143,440	140,430	231.0 ± ¹⁸ / ₁₈
1,124.2 (A)	76,330	72,690	235.0 ± ³¹ / ₃₁	1,975.3 (N)	146,860	143,370	217.0 ± ²⁰ / ₂₀
1,148.7 (Y)	78,270	74,720	230.5 ± ⁷ / ₇	1,998.0 (A)	150,330	146,340	191.0 ± ³ / ₁₀
1,175.0 (N)	80,320	76,860	219.5 ± ⁵ / ₁₀	2,025.7 (N)	154,980	150,700	200.5 ± ³ / ₁₀
1,225.7 (A)	84,220	80,900	222.5 ± ¹² / ₁₂	2,050.3 (N, N)	159,100	154,970	191.3 ± ⁸ / ₁₀
1,251.5 (N)	86,220	82,920	234.0 ± ¹² / ₁₂	2,077.5 (N)	163,670	159,690	195.5 ± ^{7.5} / _{7.5}

* (Y) indicates the presence and (N) the absence of an H₂O peak in the chromatographic analysis. Ambiguous cases are denoted (A).

into the gas chromatograph. Experiments indicate that the presence of water vapour in the analysed air is not systematic but, when it occurs, leads to measured CO₂ concentrations lower by about 13 p.p.m.v. than in its absence. For the 75 Vostok core measurements reported in this paper, we found 22 cases with a water-vapour peak and 37 cases without. The presence of this peak was ambiguous for the 16 remaining measurements, leading in those cases to an additional uncertainty. Another source of uncertainty lies in the fact that two different stainless steel containers were used for ice crushing and that the respective results differ statistically by about 5 p.p.m.v.

For each depth level our 'best estimate' of the CO₂ concentration takes into account the correction of +13 p.p.m.v. for the effect of water vapour where there were unambiguous water-vapour peaks. For consistency with previously published data obtained with the same experimental setup^{5,10,12} we added, when evaluating this 'best estimate', 5 p.p.m.v. to the CO₂ concentrations measured with the crushing container giving statistically lower values.

All the identified sources of uncertainty (experimental accuracy, effects of water vapour and crushing container) have been taken into account in plotting the CO₂ envelope shown in Fig. 1 and lead to errors around the 'best estimates' ranging from +3 to +22 p.p.m.v. on the positive side and from -3 to -16 p.p.m.v. on the negative side (Table 1). A single measurement was generally performed on each of the 66 depth levels investigated. Eight levels were nevertheless sampled and measured 2 or 3 times. For these levels the scatter of the multiple measurements is generally much smaller than the width of the envelope.

Although we are confident in the relative CO₂ variations measured, a systematic error could affect the absolute values. A recent comparison, still in progress, between results from the Physics Institute of Bern and our laboratory indicates that the

values measured in Bern are systematically higher by about 10 p.p.m.v. The results of the comparison, when completed, will be published jointly by the two groups elsewhere.

Vostok CO₂ record

Dating the air found in ice bubbles. Dating the successive ice layers at Vostok has been discussed and an ice chronology established by Lorius *et al.*¹¹. This ice chronology cannot, however, be directly applied to the dating of the CO₂ record. Indeed, owing to the air enclosure process in the ice, the air is isolated from the atmosphere in the firn layers well after the precipitation has been deposited at the surface of the ice sheet, and consequently air extracted from the ice is younger than the ice itself.

The following assumptions have been made to date the air in relation to the ice: (1) the air in the firn layers is well mixed with the atmosphere until firn pores close; (2) pore closure occurs roughly in the firn density interval 0.80–0.83 (ref. 13), that is, in the deepest layers of the firn (found between about 86 and 96 m depth at Vostok); and (3) the density–depth profile has remained unchanged over the past 160 kyr.

Based on these assumptions the difference between the age of the ice and the mean age of the air, which is dependent on accumulation rate, is evaluated taking into account the temporal changes of the accumulation rate¹¹. This calculated difference varies between about 2,500 yr for the warmest to about 4,300 yr for the coldest periods.

On the other hand, all the air enclosed in an ice sample has not been pinched off from the atmosphere at the same time but rather pore after pore over a time interval which, with the above assumptions, lies between 300 and 750 yr. This acts as a low-pass filter and means that each CO₂ measurement represents roughly an average value over several hundred years.

All the CO₂ ages given in this paper are based on the ice

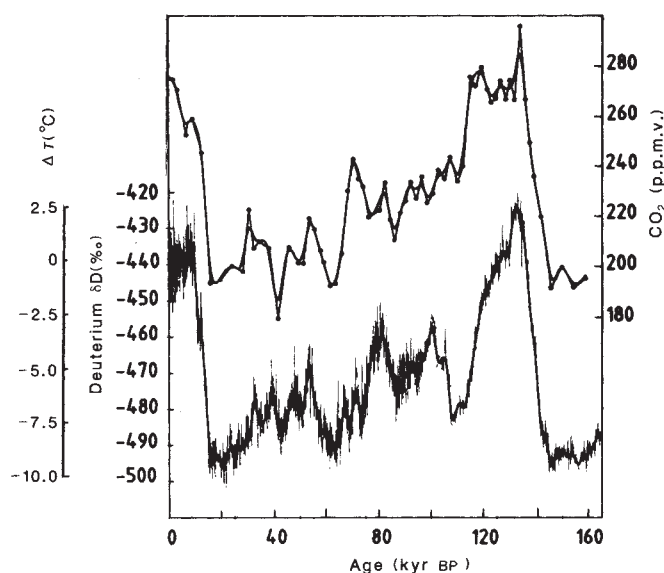


Fig. 2 CO₂ concentrations ('best estimates') and smoothed values (spline function) in p.p.m.v. plotted against age in the Vostok record (upper curves) and atmospheric temperature change derived¹⁴ from the isotopic profile (lower curve). The deuterium scale corresponds to values after correction for deuterium changes of oceanic water¹⁴.

chronology given by Lorius *et al.*¹¹, and dating of the air relative to the ice was done as described above. No better quantitative method is currently available to evaluate the age of the air. Nevertheless the two first assumptions made above probably lead to a maximum difference between the two ages (air and ice)⁵, whereas the last would lead to an age difference too small (by maybe several hundred of years) for the coldest periods.

Description. Sixty-six depth levels were investigated for CO₂ measurements along the 2,083-m core. They are spaced every ~25 m from 850 m depth to the bottom. Because of the presence of fractures in the upper part of the core, the spacing is generally larger above 850 m depth. With this sampling, the mean age difference between two neighbouring levels ranges from ~2,000 to ~4,500 yr.

The record covers the past ~160 kyr and includes the following major climatic periods: the Holocene, the last glaciation, the previous interglacial and the end of the penultimate glaciation^{11,14}. Figure 1 shows the results (best estimates and envelope obtained as described above) plotted against depth. Figure 2 shows the CO₂ variations with age of the air.

The CO₂ record exhibits two very large changes, one near the most recent part of the record (~400 m depth or 15 kyr BP) and the other near the earliest part of the record (~1,950 m depth or 140 kyr BP), between two levels centred on 190–200 and 260–280 p.p.m.v. roughly typical of the lowest and highest values of the whole record. The high level is comparable with the so-called 'pre-industrial' CO₂ level that prevailed roughly 200 years ago, just before the large anthropogenic disturbance of the CO₂ cycle; the low level ranges among the lowest values of the known geological history of carbon dioxide over the last 10⁸ yr (ref. 15). These two large CO₂ changes correspond to the transitions between full glacial conditions (low CO₂) of the last and penultimate glaciations, and the two major warm periods (high CO₂) of the record: the Holocene and the previous interglacial.

Between these two extremities of the record, that is, over the period covering mainly the last glaciation, the CO₂ shows a general decreasing trend with the lowest values found over the last part of the glaciation (between ~1,025 and 425 m depth, or 65 and 15 kyr) and with two marked breaks: a first decrease of ~40 p.p.m.v. at ~1,610 m depth (~114 kyr BP), and a second

of ~50 p.p.m.v. at ~1,070 m depth (~68 kyr BP). Other CO₂ changes are superimposed on the general decreasing trend, suggesting an oscillatory behaviour with an apparent period of ~20 kyr, especially between ~80 and 20 kyr BP. These changes are more pronounced in the last part of the glaciation, but note that some of the variability observed above 850 m depth could be due to the lower ice quality in that part of the core.

The fact that the record does not indicate a continuous long-term trend is an important result in itself regarding possible fractionation of the different gas components due to the interaction with the ice lattice and which should be reflected by such a trend. This adds confidence to the preservation of the atmospheric CO₂ record in the ice over such a long period.

Spectral analysis. The record, as shown in Fig. 2, suggests a cyclic behaviour of the CO₂ changes. Because of the potential link between CO₂ and climate, and consequently between CO₂ and astronomical cyclic forcing of climate, a spectral analysis of the CO₂ record has been done. For comparison with the δD record of Vostok (see below), the same procedure as in ref. 14 has been applied, using values equally spaced every 200 years and obtained by linear interpolation of our best estimates for CO₂ measurements. Because of the short length of the record (relative to the frequencies of interest), we combined different spectral techniques to extract reliable information. Blackman–Tukey (BT) and maximum entropy (ME) methods with and without prewhitening were applied. The BT method has a poor resolution but well defined statistical properties; the use of the ME method increases the frequency resolution but the amplitude estimate has to be interpreted with caution¹⁶. The prewhitening allows the removal of part of the variance at low frequencies and a more detailed picture in the tilt and precession frequency bands. The ME spectra shown in Fig. 3 were obtained with the Barrodale and Eriksson algorithm¹⁷.

The spectra obtained with the unprewhitened series are dominated by the 100-kyr components (116 kyr for the ME peak). The BT spectrum (not shown in Fig. 3) is relatively flat for higher frequencies but shows a significant bump for periods between 25 and 20 kyr. The ME spectrum indicates a concentration of variance in this frequency band with a well-defined peak at 21 kyr; it also displays a peak at ~55 kyr. Prewhitening, which gives more detailed information for this part of the spectrum, confirms the position of a 21-kyr peak (21.0 kyr for ME). Sensitivity analysis shows the remarkable stability of this peak ($\sigma = 0.1$ kyr) with respect to the lag value (BT) or to the autoregressive order (ME). The ME prewhitened CO₂ spectrum is thus dominated by this component of ~21 kyr and also shows a secondary peak at ~40 kyr (43.1 kyr). As the latter may result from the variance transfer of the 55-kyr peak mentioned above (unprewhitened ME spectrum), it must be interpreted very cautiously.

Because of the relatively small number of measurements with corresponding errors, any interpretation of the spectral analysis needs to be cautious. To check the significance of the spectrum results we have performed some tests with random curves placed within the error band. The results show a rather good stability of the peaks, especially for 21 kyr (20.9 ± 0.2). Some implications of this CO₂ spectrum on the CO₂–climate correlation and on the CO₂ cycle are discussed below.

Comparison with other ice records. Several ice cores from Antarctica and Greenland have already shown CO₂ variations over the past 30–40 kyr (D 10 (ref. 7), Dome C^{7,10}, Byrd⁸ Camp Century⁸, Dye 3 (refs 9, 18)). The Vostok profile is in good agreement with the general common features of these records, which show low CO₂ values (~200 p.p.m.v.) during the Last Glacial Maximum (~18 kyr BP) and an increase of the atmospheric CO₂ associated with the glacial–Holocene transition.

Other striking individual CO₂ features are also observed in some of these previous ice records. For instance, abrupt CO₂ changes are recorded in the Dye 3 core between 30 and 40 kyr BP (ref. 9) and in the Dome C core near the end of the glacial–Holocene transition^{7,10}. They do not appear in the measured

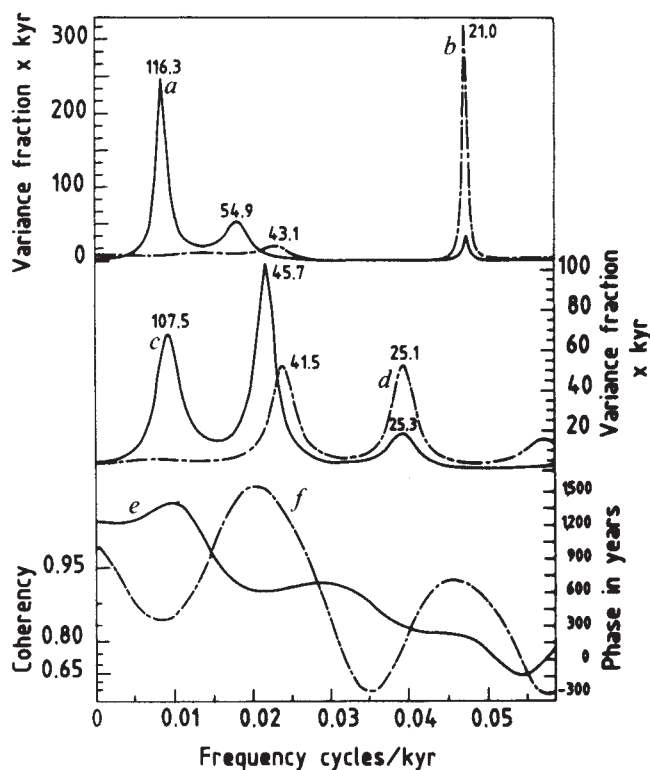


Fig. 3 *a, b*, Variance spectra of the Vostok CO₂ record. The normalized variance density is plotted against frequency in cycles per kyr. The continuous line (*a*) refers to the original series and the dashed line (*b*) to the series after prewhitening by a first-order difference filter. The spectra were obtained by the ME method applying the Barrodale and Erickson algorithm with an autoregressive order of 40%. *c, d*, Variance spectra of the Vostok isotope temperature record from ref. 14, obtained with the same procedure. *e, f*, Cross-spectral analysis between CO₂ and temperature for coherence (*e*) and phase spectrum (*f*), obtained by applying the Blackman-Tukey method. The phase (right scale) is positive when CO₂ lags the temperature.

Vostok profile, but our sampling does not allow the search for such rapid changes; furthermore, the low-pass filter effect of the air-trapping process (mentioned above) may prevent any detection in the Vostok core of the changes observed in the Dye 3 core. Note also that the high CO₂ level (>270 p.p.m.v.) observed in the Dome C core between 23 and 30 kyr BP is not found in the Vostok record. These high CO₂ values could be due to the poor quality of the Dome C ice in this part of the core¹⁹.

Thus, taking into account the resolution of the different records, the Vostok CO₂ results are consistent with the other ice records already published, covering only the most recent part of the time interval revealed by the Vostok core.

Comparison with deep-sea sediment record. $\delta^{13}\text{C}$ differences between planktonic and benthic foraminifera found in a deep-sea core (V 19-30) have been used by Shackleton *et al.*²⁰ to deduce a record of atmospheric CO₂ covering approximately the last climatic cycle and by Shackleton and Pisias²¹ to extend this record over the past 340 kyr. The basic idea, following an approach developed by Broecker²², is that a $\delta^{13}\text{C}$ difference between ocean surface and deep waters is associated with the removal of carbon from the surface by photosynthesis and consequently with the atmospheric CO₂ concentration level. The CO₂ record thus obtained agreed fairly well with the available CO₂-ice record, that is, corresponding to the period covering the past ~40 kyr.

We are now able, with the Vostok ice record, to extend the comparison over the past ~160 kyr. For this comparison we use

the more recent data by Shackleton and Pisias²¹ for the marine record. Major similarities in the general trends are observed. Both records show rapid and large increases of CO₂ concentrations associated with the two deglaciations and exhibit the highest CO₂ values during the interglacial periods. They also show together a relatively abrupt CO₂ decrease associated with the beginning of the last glaciation.

The comparison is limited in detail because of differences in timescales¹⁴ and in time resolutions between the ice and marine records. There are nevertheless significant differences between the two curves, which include the following observations.

(1) The disparity between previous interglacial and Holocene CO₂ values is much smaller in the Vostok record.

(2) The CO₂ concentrations deduced from the marine record approach Holocene levels several times during the last glaciation, unlike the Vostok record.

(3) The morphology of the Vostok CO₂ curve over the last glaciation (that is, over a large common part of the two records) is definitely not the same as that of the V 19-30 curve. Thus the general CO₂ decrease observed in the Vostok core does not appear in the marine record and the most striking change appears in the Vostok curve as a marked CO₂ decrease (between ~71 and 62 kyr BP in the Vostok timescale) and in the V 19-30 record as a large and sharp CO₂ increase (between ~60 and 55 kyr BP in the V 19-30 timescale).

(4) Spectral analysis has been done^{21,23} on the CO₂ record recovered from the marine core study. The V 19-30 spectrum exhibits peaks around all the orbital periods (~100 kyr for the Earth's eccentricity, ~41 kyr for the obliquity of the Earth's axis and 23 and 19 kyr for the precession of the equinoxes), whereas the Vostok spectrum shows as mentioned above, aside from the 110 kyr period, a peak at ~21 kyr but nothing clearly corresponding to the obliquity frequency. The comparison is, in fact, not straightforward because the record lengths (~340 kyr for V 19-30 and ~160 kyr for Vostok) as well as part of the procedures used for the spectral analysis are different. We have checked the possible influence of these differences by applying equivalent procedures to the Vostok record and to the V 19-30 data given in ref. 21, limited to the past 160 kyr. The V 19-30 spectrum still shows, in contrast with Vostok, a significant peak around the 40-kyr period.

Thus, although the CO₂ record derived from the carbon isotope differences ($\Delta\delta^{13}\text{C}$) between planktonic and benthic foraminifera found in the V 19-30 core, as well as the Vostok CO₂ record, indicate that alternations of low and high mean CO₂ concentrations and of glacial and interglacial periods are related, there are significant discrepancies between the changes observed in each of the records. Since the ice record provides *a priori* a more direct and accurate measurement of atmospheric CO₂ changes, these differences suggest that the $\Delta\delta^{13}\text{C}$ record of the V 19-30 core cannot be interpreted as a pure atmospheric CO₂ curve.

Implications for climate

In this section we compare the CO₂ profile described above with the ice isotopic record obtained on the same core¹⁴ and discuss some climatic implications. A companion paper²⁴ addresses the potential role of various forcing factors, including CO₂, for producing the temperature record derived from the isotopic composition of the Vostok ice core.

The isotopic record shows the 'dramatic' changes between the two glaciations and the two interglacial periods and provides a detailed description of other climatic variations, especially during the last glacial and the previous interglacial periods.

Our CO₂ profile and the Vostok climatic record (adapted from ref. 14) are plotted together against time in Fig. 2. The climatic significance of the isotopic profile is fully discussed in refs 11 and 14 and high (low) δD values are associated with 'warm' ('cold') temperatures. The temperature scale refers to the atmospheric surface-temperature changes in the Vostok area. The CO₂

and δD records, as obtained from the same ice core, can be directly compared after allowing for the mean age difference between the ice and the air enclosed in the ice (see section on the dating); this age difference has been taken into account in Fig. 2.

The first obvious feature when comparing the two records is the marked similarity in temporal variations. The high correlation ($r^2=0.79$) between the linearly interpolated CO_2 curve based on the best estimates and the smoothed δD curve shown in Fig. 2 reflects that generally high (low) CO_2 concentrations correspond to 'warm' ('cold') temperatures in the Vostok area. Thus the CO_2 values for the two interglacial periods (mean concentrations of 263 p.p.m.v. for the Holocene and of 272 p.p.m.v. for the previous interglacial period) are much higher than the concentrations measured in the glacial ice (on average 215 p.p.m.v.). Furthermore during the last glaciation, as mentioned above, the mean CO_2 level was higher during the first part (230 p.p.m.v. between 110 and 65 kyr BP) than during the second part (203 p.p.m.v. between 65 and 15 kyr), which also shows colder climatic conditions.

But a more detailed examination of the two records in Fig. 2 indicates differences:

(1) During the previous interglacial period and the transition towards the last glaciation, although both CO_2 and δD seem to peak simultaneously at ~ 135 kyr BP (note that the CO_2 peak is documented by only one depth level and therefore needs to be confirmed), the CO_2 record shows relatively high and constant values over at least the 10 kyr after the peak (see Fig. 2), whereas the δD curve clearly decreases over the same time interval. A similar effect, although less obvious, seems to appear around 75 kyr BP when the δD curve shows a decreasing trend whereas the average CO_2 apparently remains constant.

(2) The first 'isotopically cold' peak observed during the last glaciation, around 110 kyr BP, has no low- CO_2 companion peak.

The superposition of the CO_2 and δD records also provides information on the comparative timing of the most significant changes in the two signals. The sparse CO_2 sampling and to a smaller extent the uncertainties in the CO_2 measurements and on the age difference between the air trapped in the ice and the ice itself make the determination of a time difference between the two signals of Fig. 2 only possible for differences larger than 3500 yr. Within this uncertainty, the following trends are observed: (1) δD and CO_2 changes appear almost simultaneously when proceeding from glacial to interglacial conditions; (2) as noted above, the situation could be different when proceeding from the previous interglacial to the last glaciation with the CO_2 change clearly lagging behind the δD decrease.

Furthermore information on the comparison between the two time series can be inferred from the spectral analysis. As mentioned above, the CO_2 spectrum reveals a well-defined peak at 21 kyr and, with prewhitening, a relatively weak secondary peak at ~ 40 kyr that must be interpreted very cautiously. This contrasts (see Fig. 3) with the temperature spectrum deduced from the δD record and given by Jouzel *et al.*¹⁴. The temperature spectrum does indeed indicate peaks at ~ 40 and ~ 20 kyr (that is, which can be associated with the obliquity and precession cycles), with the 40-kyr peak dominating the 20-kyr. The results (Fig. 3) of the cross-spectral analysis for coherence and phase spectrum between CO_2 and temperature (δD) show high coherence in the orbital frequency band and indicate that CO_2 could lag the temperature slightly. This lag is not statistically significant but could reflect, as mentioned above, the temperature signal decreasing before the CO_2 when proceeding from the interglacial to the glacial period.

The above comparison deals with a CO_2 signal that should be of global atmospheric significance and a climatic signal that is *a priori* relevant to Antarctica and consequently to more local conditions. The Vostok δD record shows, in fact, important climatic changes on a worldwide scale^{11,14}. In this more general context it is interesting to compare the CO_2 level measured over

the previous interglacial with the present atmospheric CO_2 concentrations, because this period of the past has been proposed as an analogue for a future CO_2 -warmed world. The Vostok results indicate a mean CO_2 concentration (272 p.p.m.v.) comparable with the 'pre-industrial' CO_2 level (estimated to be 275 ± 10 p.p.m.v. (ref. 25)). Furthermore, the CO_2 peak of about 300 p.p.m.v. found at ~ 135 kyr BP, if confirmed, is significantly lower than the current level of about 345 p.p.m.v.

The high correlation between atmospheric CO_2 and isotopic temperature found in the Vostok record suggests that the radiative effect of CO_2 coupled with associated feedback mechanisms appears to be a serious candidate as one of the driving forces behind the palaeo-temperature changes recorded in the Vostok core. The relative weights of different types of forcing in the reconstruction of the Vostok temperature variations are discussed in the companion paper²⁴. This approach points out the probable importance of CO_2 in influencing the main feature of the climate during the last climatic cycle.

Examining the parts of the CO_2 and δD records containing the most pronounced changes, although the two transitions from glacial to interglacial conditions shown by the record indicate almost simultaneous changes for CO_2 and isotopic temperature, the situation is different when looking at the transition between the previous interglacial period and the last glaciation. As shown in Fig. 2, a large portion, about 7 °C, of the isotopic temperature decrease leading to the glacial conditions takes place between ~ 132 and 115 kyr, a time during which no significant CO_2 changes are observed. Our results thus indicate that climatic forcings other than CO_2 induced a large part of the temperature change over Antarctica linked with the interglacial-glacial transition. The marked CO_2 decrease (from ~ 275 to 235 p.p.m.v.) appears only during the final phase of the climatic cooling (between ~ 115 and 110.5 kyr) and is associated with a decrease of ~ 2 °C in the Vostok isotopic temperature.

Based on the above discussion, we suggest that the interactions between climate and CO_2 could be different depending on whether the climate shifts from glacial to interglacial or from interglacial to glacial conditions. In the latter case, orbital and orbitally derived forcing could have influenced the marked cooling trend in the Antarctic with a dominant contribution to the temperature decrease, before the potential climatic effect of the change in CO_2 concentrations. Note that the second (in terms of amplitude) cooling trend, found around 75 kyr BP in the Vostok record, exhibits a similar behaviour of the CO_2 - δD comparative variation.

Implications for CO_2 cycle

As well as providing information on the CO_2 -climate interaction, our CO_2 profile allows us to discuss the modifications of the carbon cycle during the last climatic cycle. Based on previous ice core results over the past 30–40 kyr, several models have been proposed to explain the increase of atmospheric CO_2 associated with the last deglaciation. Detailed descriptions of these models can be found in Broecker and Peng²⁶ or Berger and Keir²⁷. These models are based on the idea that the biological productivity of the ocean controls the atmospheric CO_2 but they differ in the mechanisms driving these productivity variations.

In one family of models, sea-level variations are responsible for the atmospheric CO_2 changes^{22,28}. The general trend of the Vostok CO_2 profile is similar to the benthic $\delta^{18}O$ records, which mainly reflect ice volume or sea level (for example see Fig. 2b, *e* of the companion paper²⁴), making these models attractive. Berger and Keir²⁷ and Keir and Berger²⁹ attempt to model the past atmospheric CO_2 modifications due to sea level variations. Choosing appropriate parameters, the morphology of the derived CO_2 curve is close to that of the Vostok curve and the CO_2 variations lag the sea-level changes by ~ 1 kyr. To test this time lag, it is necessary first to evaluate the phase relationship between the CO_2 and the Vostok temperature and secondly

between the Vostok temperature and the sea level. In deglaciations, as already mentioned, the CO_2 begins to increase almost simultaneously with the Vostok temperature. The precise timing between the temperature and the sea level is difficult to obtain because these two signals are not recorded in the same medium. It is nevertheless reasonable to assume that the sea level cannot increase before the Antarctic temperature, as suggested by the interpretation of the marine sediment record, which indicates that changes in Southern Hemisphere sea-surface temperature have generally led the changes in ice volume^{30,32}. Although denser sampling would be required for a more definite conclusion, the Vostok results thus suggest that at the end of glaciations the CO_2 increased before the sea level and not after as predicted by the models mentioned. At the beginning of the last glaciation (~120 kyr BP), the CO_2 decreased ~10 kyr after the temperature and thus most probably after the sea-level change. Although this phase relation cannot exclude the influence of the sea level on the CO_2 variations during this period, it is important to note that the CO_2 decrease is very sharp (35 p.p.m.v. in 2 kyr) and seems incompatible with the mechanisms involved in the models proposed. We suggest that this abrupt change of the atmospheric CO_2 between two nearly constant values could be due to an abrupt modification of the deep oceanic circulation possibly linked with the sea-level decrease. Such rather rapid circulation changes associated with climatic transitions have been shown to occur³³.

In another family of models, changes in oceanic circulation affect the biological productivity of the surface ocean and thus the atmospheric CO_2 concentration. These models stress changes occurring in the high latitude deep-water formation areas³⁴⁻³⁶ and in the upwelling zones at low latitudes³⁵. The spectral analysis of the Vostok CO_2 record shows, as well as the 110-kyr component, a very weak and doubtful 41-kyr and a clear 21-kyr peak. This feature can give some indications of the latitudes involved in the forcing of the atmospheric CO_2 . The latitudinal variation of the insolation spectra is complex. Although the presence of an obliquity component (41 kyr) seems to be a characteristic of high latitudes, this component is negligible around the equinoxes and becomes important around the solstices³⁷. From marine proxy data, the spectral analysis of the sea-surface temperature of the North Atlantic shows that, apart from the 110-kyr component, the 41-kyr component is dominant for latitudes greater than 45° N (ref. 38), and we can note that the Vostok temperature profile also exhibits a strong 41-kyr period¹⁴. The quasi-absence of a 41-kyr period in the CO_2 record could thus suggest that the high latitudes do not have a major influence on the CO_2 variations. Nevertheless we cannot exclude an influence of these latitudes during part of the year, in particular around the equinoxes.

At low latitudes, upwelling areas are known to play a key role in the total biological productivity of the oceans and may thus influence the atmospheric CO_2 significantly. Evidence exists from the reconstructed sea-surface temperatures that upwellings were stronger during the Last Glacial Maximum (18 kyr BP)³⁹. Similar conclusions are obtained from the palaeoreconstructions of trade-wind strength⁴⁰⁻⁴², responsible for the upwelling currents. Some of these indicators exhibit similarities with the atmospheric CO_2 , for instance at the end of the two last glaciations the upwelling productivity off the NW African coast decreased before the increase in sea level⁴⁰, as did the winter trade-wind speed near the Peru coast⁴¹. Note also that the upwelling productivity off NW Africa was high during only the second part of the glaciation, that is, between ~70 and 15 kyr BP⁴¹, and that the Vostok aluminium concentration profile exhibits three very pronounced spikes suggesting stronger atmospheric circulation at ~150, ~70 and ~20 kyr BP⁴³. These spikes correspond to very low CO_2 values (~200 p.p.m.v.).

All these facts suggest that atmospheric circulation, coupled with oceanic circulation, at least through the upwelling currents, could have played an important role in past CO_2 variations.

This influence may have been especially important between 70 and 15 kyr BP and thus could be responsible for at least part of the lower CO_2 values observed during this period compared with the values recorded during the previous part of the glaciation (110–70 kyr BP).

In summary, it seems that the atmospheric CO_2 variations recorded in the Vostok core could be explained by two kinds of changes in the oceanic circulation, that is, the deep changes possibly driven by sea level and the surface changes driven by atmospheric circulation. In this case only deep circulation changes could be the cause of the CO_2 variations between the last interglacial and the first part of the glaciation (110–70 kyr), whereas both deep and surface circulation could be responsible for the lower CO_2 values found during the second part of the glaciation (70–15 kyr) relative to the first part. During glacial-interglacial transitions, the CO_2 increase could be initiated by surface circulation changes relayed by the deep circulation changes associated with the sea-level increase.

Conclusions

An atmospheric CO_2 record over the past 160 kyr has been obtained from the Vostok ice core. This CO_2 record is probably the purest available, covering the last climatic cycle. The CO_2 changes thus revealed, which are of global significance, are well correlated with the Antarctic temperature record derived from the ice isotopic profile measured on the same core. Such a high correlation would be expected if CO_2 plays an important role in forcing the climate.

The results also suggest a different behaviour of the relative timing between CO_2 and Antarctic climate changes depending on whether we proceed from a glacial to an interglacial period or vice versa. This may have implications concerning the relation between cause and effect inside the CO_2 -climate system, but more detailed measurements in the transition parts of the record are required before firm conclusions can be drawn. Long-term CO_2 changes are dominated by marked glacial-interglacial oscillations between ~190–200 and 260–280 p.p.m.v. A period of ~20 kyr similar to the orbital precession period, appears in the decreasing CO_2 trend covering most of the last glaciation and is supported by spectral analysis. These CO_2 changes could be linked with oceanic circulation changes but a more precise interpretation would in particular require the determination of the $\delta^{13}\text{C}$ of the CO_2 extracted from the air bubbles trapped in the ice and a comparable chronology for marine and ice core records.

We thank all Soviet and French participants in ice drilling and sampling. We acknowledge the efficient logistic support of the Soviet Antarctic Expeditions, the US National Science Foundation (Division of Polar Programs) and the French Polar Expeditions. We thank C. Genthon and J. Jouzel for performing the CO_2 spectral analysis and J. C. Duplessy, C. Genthon and J. Jouzel for helpful discussions. This work was supported in France by the Commission of the European Communities (Climatology Research Programme), CNRS/PIREN, TAAF and in the Soviet Union by Soviet Antarctic Expeditions.

Received 26 March; accepted 17 July 1987.

1. Lorius, C. & Raynaud, D. in *Carbon Dioxide: Current Views and Developments in Energy/Climate Research* (eds Bach, W. et al.) 145–176 (Reidel, Dordrecht, 1983).
2. Stauffer, B. & Oeschger, H. *Ann. Glaciol.* 7, 54–59 (1985).
3. Raynaud, D., Delmas, R., Ascencio, J. M. & Legrand, M. *Ann. Glaciol.* 3, 265–268 (1982).
4. Neftel, A., Moor, E., Oeschger, H. & Stauffer, B. *Nature* 315, 45–47 (1985).
5. Raynaud, D. & Barnola, J. M. *Nature* 315, 309–311 (1985).
6. Pearnan, G. I., Etheridge, D., De Silva, F. & Fraser, P. J. *Nature* 320, 248–250 (1986).
7. Delmas, R. J., Ascencio, J. M. & Legrand, M. *Nature* 284, 155–157 (1980).
8. Neftel, A., Oeschger, H., Schwander, J., Stauffer, B. & Zumburn, R. *Nature* 295, 220–223 (1982).
9. Stauffer, B., Hofer, H., Oeschger, H., Schwander, J. & Siegenthaler, U. *Ann. Glaciol.* 5, 160–164 (1984).
10. Raynaud, D. & Barnola, J. M. in *Current Issues in Climate Research* (eds Ghazi, A. & Fantechi, R.) 240–246 (Reidel, Dordrecht, 1985).
11. Lorius, C. et al. *Nature* 316, 591–596 (1985).
12. Barnola, J. M., Raynaud, D., Neftel, A. & Oeschger, H. *Nature* 303, 410–412 (1983).

13. Schwander, J. & Stauffer, B. *Nature* **311**, 45–47 (1984).
14. Jouzel, J. *et al.* *Nature* **329**, 403–408 (1987).
15. Gammon, R. H., Sundquist, E. T. & Fraser, P. J. in *Atmospheric Carbon Dioxide and the Global Carbon Cycle* (DOE Report ER-0239) 25–62 (1985).
16. Pestiaux, P. & Berger, A. L. in *Milankovitch and Climate* (eds Berger, A. *et al.*) 417–445 (Reidel, Dordrecht, 1984).
17. Barrodale, I., & Erickson, R. E. *Geophys* **45**, 420–432 (1980).
18. Stauffer, B., Neftel, A., Oeschger, H. & Schwander, J. in *Greenland ice core: Geophysics, Geochemistry and the Environment* (eds Langway, C. C. *et al.*) (Geophysical Monograph 33) 85–89 (American Geophysical Union, Washington DC, 1985).
19. Barnola, J. M. thesis, Université Scientifique et Médicale de Grenoble (1984).
20. Shackleton, N. J., Hall, M. A., Line, J. & Cag, Shuxi *Nature* **306**, 319–322 (1983).
21. Shackleton, N. J. & Pisias, N. G. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) (Geophysical Monograph 32) 303–317 (American Geophysical Union, Washington, DC, 1985).
22. Broecker, W. S. *Progr. Oceanogr.* **11**, 151–197 (1982).
23. Pisias, N. G. & Shackleton, N. J. *Nature* **310**, 757–759 (1984).
24. Genthon, C. *et al.* *Nature* **329**, 414–418 (1987).
25. Scientific Committee on Problems of the Environment in *Scope 29: The Greenhouse Effect, Climatic Change and Ecosystems* (eds Bolin, B. *et al.*) xxv–xxxii (Wiley, Chichester, 1986).
26. Broecker, W. S. & Peng, T. H. *Radiocarbon* **28**, 309–327 (1986).
27. Berger, W. H. & Keir, R. S. in *Climate Processes and climate sensitivity* (eds Hansen, J. E. & Takahashi, T.) (Geophysical Monograph 29) 337–350 (American Geophysical Union, Washington, DC, 1984).
28. Berger, W. H. *Naturwissenschaften* **69**, 87–88 (1982).
29. Keir, R. S. & Berger, W. H. *J. geophys. Res.* **88**, 6027–6038 (1983).
30. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
31. CLIMAP project members *Quat. Res.* **21**, 123–224 (1984).
32. Labeyrie, L. D. *et al.* *Nature* **322**, 701–706 (1986).
33. Duplessy, J. C. & Shackleton, N. J. *Nature* **316**, 500–507 (1985).
34. Knox, F. & McElroy, M. B. *J. geophys. Res.* **89**, 4629–4637 (1984).
35. Siegenthaler, U. & Wenk, T. *Nature* **308**, 624–626 (1984).
36. Sarmiento, J. L. & Toggweiler, J. R. *Nature* **308**, 621–624 (1984).
37. Berger, A. & Pestiaux, P. in *Milankovitch and Climate* (eds Berger, A. *et al.*) 83–111 (Reidel, Dordrecht, 1984).
38. Ruddiman, W. F. & McIntyre, A. *Bull. geol. Soc. Am.* **95**, 381–391 (1984).
39. CLIMAP project members *Science* **191**, 1131–1136 (1976).
40. Sarntheim, M., Winn, K. & Zahn, R. in *Biviers Symposium on Abrupt Climatic Changes* (in the press).
41. Molina Cruz, A. *Quat. Res.* **8**, 324–328 (1977).
42. Boyle, E. J. *J. geophys. Res.* **88**, 7667–7680 (1983).
43. DeAngelis, M., Barkov, N. I. & Petrov, V. N. *Nature* **325**, 318–321 (1987).

Vostok ice core: climatic response to CO₂ and orbital forcing changes over the last climatic cycle

C. Genthon*, J. M. Barnola†, D. Raynaud†, C. Lorius†, J. Jouzel*, N. I. Barkov‡, Y. S. Korotkevich‡ & V. M. Kotlyakov§

* Laboratoire de Géochimie Isotopique—LODYC (UA CNRS 1206), CEA/IRDI/DESICP/DPC, 91191 Gif sur Yvette Cedex, France

† Laboratoire de Glaciologie et Géophysique de l'Environnement, BP 96, 38402 Saint Martin d'Hères Cedex, France

‡ Arctic and Antarctic Research Institute, Beringa Street 38, 199226 Leningrad, USSR

§ Institute of Geography, Academy of Sciences of the USSR, 29 Staronomety, Moscow 109107, USSR

Vostok climate and CO₂ records suggest that CO₂ changes have had an important climatic role during the late Pleistocene in amplifying the relatively weak orbital forcing. The existence of the 100-kyr cycle and the synchronism between Northern and Southern Hemisphere climates may have their origin in the large glacial-interglacial CO₂ changes.

FROM the Vostok Antarctic ice core drilled by the Soviet Antarctic Expeditions, we have, in two companion articles^{1,2}, derived time series of the central East Antarctic atmosphere temperature and of atmospheric CO₂ content spanning the past 160 kyr. These studies suggest that over this period the climate at Vostok was, at least in part, driven both by the variations in atmospheric CO₂ and by orbital forcing which would be exerted through insolation changes in the Northern Hemisphere and locally³. Here we examine the relative influence of these various forcings on a quantitative basis by calculating the temperature signal as a response to CO₂ and forcings in the Northern Hemisphere and locally. Our simple model can explain >90% of the climate variance at Vostok. It suggests that CO₂ plays an important role in influencing the Earth's climate during the late Pleistocene.

For two decades the Milankovitch hypothesis⁴ that the great Northern Hemisphere ice sheets are controlled by the seasonal contrast of insolation at ~65°N has received support, particularly from a thorough analysis of the δ¹⁸O deep-sea record^{5,6}. But the implied large amplification of this relatively weak forcing (the total insolation received by the planet varied by <0.6% over the past 10⁶ yr), the observed dominant 100-kyr cycle and the synchronism between the termination of major glaciations in the Northern and Southern Hemispheres do not fit this hypothesis well^{7–12}. Further developments of the theory of Pleistocene climates have included the nonlinear response of ice sheets to this orbital forcing^{8,13–17} and the role of oceanic circulation^{11,18,19}, ice-shelf destruction^{20,21} and sea-level changes^{9,21}. More recently the glacial-interglacial CO₂ changes shown from measurements in air trapped in ancient ice^{22,23} and later indirectly derived from δ¹³C deep-sea core data^{24,25}, have focused attention on the possible contribution of the CO₂ greenhouse effect to these major climatic changes^{10–12,17,25–28}.

To address the problem of glacial-interglacial climatic

changes, the Vostok ice-core results are important because they give access in the same core to: (1) a continuous isotope temperature series which is, at least qualitatively, of relatively large geographical significance¹ and (2) a direct estimate of the atmospheric CO₂ content which is global in character². These two series are dominated by a clear ~100 kyr signal. They show concentration of variance near orbital frequencies and suggest that there is an interaction between CO₂, orbital forcing and climate on the timescale of glacial-interglacial episodes. Because all the parts of the climatic system (atmosphere, ocean, cryosphere, biosphere) are involved, such interactions are complex. Considering this complexity, then, we shall limit our present purpose to giving and discussing the results of a multivariate analysis between a single climatic output (the Vostok isotope temperature record, Δ*T*_s, shown in Fig. 1a) and multiple climatic inputs. In essence we perform a statistical analysis without specifically accounting for the physical mechanisms involved. We consider three types of climatic forcings: (1) a Southern Hemisphere input, (2) a Northern Hemisphere input, and (3) a CO₂ input.

Methods

These climatic forcings were chosen to cover as far as possible the full range of hypotheses currently proposed in this field. Figure 1 shows corresponding time series, the insolation curves being calculated using the Berger algorithms²⁹.

As a Southern Hemisphere (SH) input we used either the local insolation (78°S) received during the entire year (*s*₁) or the mean daily November insolation at 60°S (*s*₂). As mentioned in ref. 3, *s*₁ shows visual similarities to the Vostok isotope record, which suggests there is a link between Δ*T*_s and this annual insolation; *s*₂ is thought to play an important role in the sea-ice extent around Antarctica³⁰. In all of the cases considered, the

Vostok temperature is assumed to be in phase with the SH forcing. Additional tests were made using the mid-December insolation at 78° S; the results are quite similar to those produced with s_2 . We also attempted to use the solar radiation absorbed at the snow surface ($s_1(1-\alpha)$, where α is the albedo) accounting for the dependence of α on the solar elevation³¹. Although this effect is relatively important at high latitudes, on a seasonal basis, there is no significant modification of the shape of the local forcing in the long term. These two additional approaches were not explored further.

The Northern Hemisphere (NH) component was taken to be either the July insolation at 65° N (n_1), the Imbrie and Imbrie⁸ ice volume theoretically derived from this insolation curve (n_2), or the $\delta^{18}\text{O}$ reconstructed ice volume (n_3) of Martinson *et al.*³². n_1 and n_2 are alternative ways of examining the Milankovitch hypothesis; n_3 is a more direct estimate of the ice volume change. The ice volume change is important as it may be the link between the climates in the two hemispheres, for example through oceanic circulation or sea-level changes. This NH forcing is not expected to be in phase with the SH climatic response^{9,10}. We introduced an open parameter, $\Delta\phi$, representing the shift between the NH input and the Vostok climatic output (equation (2)). There is no significant influence of $\Delta\phi$ on the result of the multivariate analysis. All the results given in Table 1 correspond to $\Delta\phi = 0$.

Our implicit assumption about the CO_2 input is that the changes in the CO_2 content of the atmosphere are part of climatic forcing. Two key ideas to keep in mind in the final assessment of the results of our statistical analysis are that CO_2 changes might just be a consequence of climatic change without much effect on the climatic change itself and are possibly triggered by orbital forcing, as suggested by the presence of a precessional component in the CO_2 record².

The radiative forcing associated with the CO_2 variations was evaluated in terms of its direct temperature effect (without any climatic feedback) from the empirical formula computed by Hansen *et al.*³³ which fits, to better than 1%, the results of a radiation scheme of a three-dimensional model³⁴ for CO_2 between 100 and 1,000 p.p.m.v. With respect to the Holocene temperature, it yields

$$\Delta T_{\text{CO}_2} = \ln(1 + 1.2X + 0.005X^2 + 1.4 \times 10^{-6}X^3) - 6.64 \quad (1)$$

X being the atmospheric CO_2 concentration in p.p.m.v. (as recorded in the Vostok core). This calculated temperature change without feedbacks, ΔT_{CO_2} , is given in Fig. 1b, showing a total amplitude of $\sim 0.6^\circ\text{C}$. Owing to oceanic thermal inertia, there is a delay in the climatic response to a CO_2 change but by no more than $\sim 100\text{ yr}$ ^{34,35}. It is not necessary to account for this delay, which is less than the period over which the atmos-

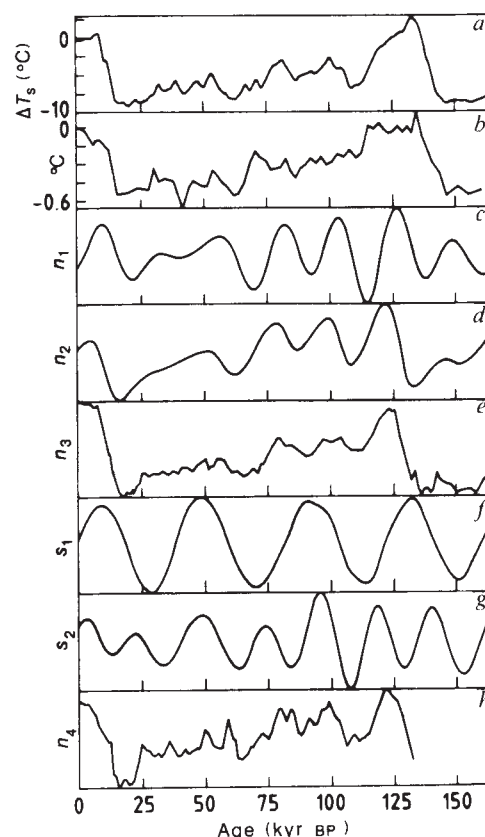


Fig. 1 Time series of the climatic inputs and of the output in the Vostok record. a, Vostok isotope temperature change, ΔT_s (Lorius *et al.*³ chronology); b, Vostok atmospheric CO_2 expressed as direct radiative forcing; c, insolation, July 65° N; d, ice volume change from the Imbrie and Imbrie model; e, estimated ice volume change from deep sea core $\delta^{18}\text{O}$; f, total insolation at 78° S during the entire year; g, insolation, November, 60° S; h, estimated ice-volume change from Labeyrie *et al.*³⁸. Curves c-h are scaled so as to have about the same amplitude.

pheric CO_2 signal is smoothed by the air-trapping process in the ice and far less than the average CO_2 sampling interval².

We applied the numerical procedure described in Jenkins and Watts³⁶ writing the climatic output, ΔT_s , as

$$\Delta T_s(t) - \overline{\Delta T_s} = a_{\text{CO}_2}(\Delta T_{\text{CO}_2}(t) - \overline{\Delta T_{\text{CO}_2}}) + b(s_i(t) - \overline{s_i}) + c(n_i(t - \phi) - \overline{n_i}) + z(t) \quad (2)$$

Table 1 Results of the multivariate analysis

	Case	Output	SH	NH	r^2	a_{CO_2}	$w_{\text{CO}_2}(\%)$	$w_{\text{SH}}(\%)$	$w_{\text{NH}}(\%)$
Lorius <i>et al.</i> ³ chronology	1	ΔT_s	s_1	n_1	0.90	13.5	78	16	6
	2	ΔT_s	s_1	n_2	0.88	13.3	79	20	1
	3	ΔT_s	s_1	n_3	0.89	12.7	75	20	5
	4	ΔT_s	s_2	n_1	0.86	14.0	85	3	12
Redated Vostok series	5	ΔT_{s1}	s_1	n_1	0.90	13.7	79	8	13
	6	ΔT_{s1}	s_1	n_2	0.87	12.5	75	13	12
	7	ΔT_{s1}	s_1	n_3	0.90	9.4	55	11	34
	8	ΔT_{s2}	s_1	n_3	0.92	6.2	35	3	62
0-110-kyr period	9	ΔT_{s3}	s_1	n_1	0.85	12.7	75	13	12
	10	ΔT_{s3}	s_1	n_2	0.84	11.9	71	16	13
	11	ΔT_{s3}	s_1	n_3	0.92	4.9	27	9	64
	12	ΔT_{s3}	s_1	n_4	0.84	11.0	66	15	19

The climatic output at Vostok (temperature change) is given in column 2 and the chosen Southern and Northern Hemisphere inputs in columns 3 and 4 (see text); r^2 is the square of the correlation between ΔT (the calculated temperature change at Vostok) and the output at Vostok; a_{CO_2} is the calculated amplification of the CO_2 temperature effect. The relative weights, w , of each contribution (columns 7-9) represent the percentage of variance explained by each of the three inputs.

where t is the time, the variables with an overbar represent the means of each time series, and s_i and n_i are the SH and NH inputs discussed above. The first three terms on the right of equation (2) correspond to the temperature change that may be attributed to each of the three inputs, and $z(t)$ is a noise term. This equation is solved in such a way to minimize $z(t)$ in a least-square sense using resampled series with equally spaced points every 100 yr. Coarser sampling or the use of slightly smoothed curves for data series were examined but they do not affect the results significantly. The best estimate of ΔT_s , from the three climatic inputs, $\Delta T_s(t) - z(t)$, will be referred to as the 'calculated' Vostok temperature change and denoted by ΔT .

The term a_{CO_2} is the ratio of the temperature change that may be attributed to CO_2 , to the direct radiative effect of CO_2 changes (without feedback); it will be referred to as the CO_2 amplification factor. This definition parallels that of the climate feedback factor, f , evaluated from climate model simulations as the ratio of the overall predicted temperature change to the ΔT_{CO_2} without feedbacks. Note that the precision of the determination of a_{CO_2} is no better than that of ΔT_s , that is, $\sim 20\%$ as discussed in ref. 1. Comparison of the amplification (multivariate analysis) and feedback (climate modelling) factors will be useful in the discussion of the CO_2 climate contribution; such a comparison is not greatly affected by this 20% uncertainty. Unlike a_{CO_2} , b and c are not directly expressed in term of temperature changes.

This linear approach is admittedly simplistic because there are certainly nonlinear interactions between the different parts of the climatic system^{8,14-17}. Owing to our strategy not to attach the Vostok dating to other palaeoclimatic or forcing series, the use of a more elaborate gain and phase model^{36,37} is not justified. However, the influence of various redatings for the period before 110 kyr BP, over which the Vostok and marine chronologies noticeably disagree¹, was also examined. We see that despite its simplicity, this approach provides interesting results about the relative contributions of various forcings. These results are discussed with regard to the 12 cases listed in Table 1. Cases 1-4 refer to the Lorius *et al.*³ chronology, cases 5-8 to redated Vostok series as described below, and cases 9-12 to the 0-110-kyr period.

Results

Cases 1-4. The first three cases test the NH forcings, keeping the SH input fixed (78° S annual insolation). The temperature output and the CO_2 forcing are those of Fig. 1a, b with the Lorius *et al.*³ Vostok chronology based on a glaciological model. The best correlation ($r^2 = 0.90$) is obtained with the July insolation at 65° N. The use of an ice-volume forcing, either calculated (n_2) or reconstructed (n_3) does not improve this correlation. In cases 1-3 the SH contribution (s_1) is significant (up to 20%). This contribution practically disappears if another SH input is used, either s_2 (November, 60° S) as illustrated by case 4 or the mid-December local insolation (not listed). The common feature of all these analyses is that the main contribution comes from CO_2 with a very high associated amplification (~ 13). This is not surprising as the ΔT_{CO_2} input is already well correlated with ΔT_s ($r^2 = 0.78$). There is, however, an interesting contribution of the other forcings as illustrated in Fig. 2, which shows ΔT and its spectrum along with the same curves for ΔT_s . The NH forcing (July, 65° N) clearly participates in simulating the 110-kyr BP temperature minimum, and the SH forcing (local forcing) brings power in the ~ 40 -kyr frequency band, two important features of the ΔT_s record missing in the CO_2 input^{1,2}.

Cases 5-8. One possible reason for the relatively low contribution of the orbital input to ΔT (the sum of NH and SH contributions never exceeds 25% in cases 1-4) is that the Vostok core is dated independently of such forcings. On the other hand, there is no significant uncertainty in the relative timing of ΔT_s and CO_2 changes. But we believe that the duration of the last interglacial period was about the same in the Vostok atmospheric record as in the marine records¹. An inevitable consequence is that one, at least, of the Vostok and deep-sea core datings has

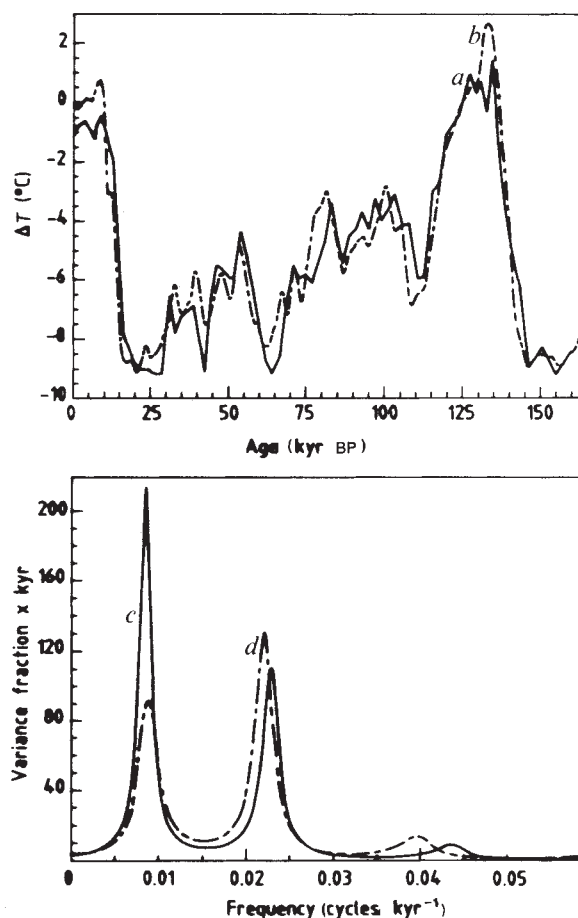


Fig. 2 Top, the calculated temperature (ΔT) obtained with the climatic inputs of case 1 (a, continuous line) plotted with the Vostok isotope temperature of ref. 1 (b, ΔT_s , broken line). Bottom, the variance spectra of these two curves. The normalized variance density is plotted against frequency in cycles per kyr on linear scales: c, ΔT ; d, ΔT_s . These spectra were obtained by the Maximum Entropy method, applying the Barrodale and Ericksson⁴² algorithm with an autoregressive order of 40%.

to be modified before ~ 110 kyr BP. We examined the influence of the chronology on our multivariate analysis in testing various redatings.

Cases 5-7 refer to a Vostok timescale modified in such a way as to put the Vostok temperature roughly in phase with that of the sea surface around Antarctica (ΔT_{s1}) taken from RC11-120 (ref. 32). The CO_2 amplification does not significantly decrease ($a_{\text{CO}_2} \approx 13$) with either July at 65° N (case 5) or Imbrie ice volume (case 6) as NH input. This is not surprising as $< 35\%$ of the temperature variance is directly explained by this type of NH forcing. On the other hand, this amplification drops with the reconstructed ice volume (case 7) down to ~ 9 .

This lowering is further accentuated if the ice volume is put in phase with the Vostok record before 110 kyr BP (ΔT_{s2} , case 8, $a_{\text{CO}_2} = 6.2$). In case 8, the explained variance is also > 0.90 ($r^2 = 0.92$) but now with a predominant contribution of the NH forcing (63% compared with 35% for CO_2), reflecting the very high correlation ($r^2 = 0.88$) observed between the deep-sea core $\delta^{18}\text{O}$ curve (n_3) and the redated Vostok temperature (ΔT_{s2}). Such a high correlation could at least partly result from the fact that the ice volume and the Antarctic temperature respond in the same way to the same common forcing and not from a direct influence of ice-volume change on the Vostok temperature. This would artificially increase the contribution of the NH component in our analysis. Consequently, the CO_2 amplification inferred from case 8 should be considered a lower limit for this parameter.

This general picture given by cases 5-8 is not greatly affected

if a slightly different Vostok timescale is used (for the period before 110 kyr BP) or if it is the deep-sea record that is fitted in phase with the Vostok series as dated by Lorius *et al.*³.

Cases 9–12. The last set of results deal with the period since 110 kyr BP (ΔT_{83}). This shortening of the records lowers the accuracy of the analysis but has the advantage of eliminating many of the problems created by the apparent dating discrepancy between the ice and deep-sea records. This approach essentially supports the results obtained with redated series, but with a further lowering of the CO₂ amplification ($a_{\text{CO}_2}=4.9$) with the Martinson *et al.* series³² as the NH input (case 11). There is still a relatively low NH contribution (and high CO₂ amplifications) with orbital forcing (cases 9 and 10).

For this period, we also used the ice-volume change recently estimated by Labeyrie *et al.*³⁸ (Fig. 1h, n_4). This results (case 12) in a relatively low influence of this NH input (19%) and intermediate CO₂ amplification (11.0). The difference from case 11 reflects the lower ice-volume–Vostok-temperature correlation obtained with this estimate than with that of Martinson *et al.* ($r^2=0.59$ instead of 0.88). It indirectly confirms that part of the Martinson *et al.* signal could originate with temperature, explaining its high correlation with the Vostok record.

Discussion

Again we are aware of the limitations of our multivariate analysis, some of them emerging from the above discussion. First, it may be useful to explore a wider range of possible forcings and associated feedbacks. For example, one can consider, the introduction of different insolation curves or of different model-derived^{14,15} or reconstructed ice volume. Effects such as the change in the atmospheric optical depth resulting from variations in aerosol loading³⁹ or in the concentration of trace gases⁴⁰ may affect the climate. We shall note that such parameters are⁴¹ or will be determined from the Vostok ice. In addition it is very important to get accurate absolute dating of the palaeoclimatic series or, at least, to correlate the deep-sea and ice records directly¹.

On the other hand, within these limits, our study allows us to suggest some conclusions about the climatic contribution of various forcings as inferred from our statistical analysis. They can be summed up as follows.

(1) The SH influence is maximum (20%) if assumed to be linked with the local annual insolation; this influence decreases if the record is redated before 110 kyr BP (13% in case 6).

(2) The NH contribution is different if exerted through an orbital (n_1) or orbitally derived (n_2) input or through the $\delta^{18}\text{O}$ ice-volume record. Only in this last case (n_3) does the NH relative contribution become predominant ($\leq 64\%$).

(3) The relative contribution of CO₂ is generally $>50\%$ and as high as 84%. CO₂ contributions $<50\%$ are, however, observed when the reconstructed ice volume of Martinson *et al.* is used as an NH input (cases 8 and 11).

In the remaining part of this discussion we focus on how these results affect our current understanding of the Earth's climatic system.

The influence of the local annual insolation was examined in ref. 3. The relative range of variation of this parameter is $\sim 7\%$, which would have, through direct radiative forcing^{34,43}, an effect of $\sim 4^\circ\text{C}$ on the local equilibrium temperature ($\Delta T_e/T_e = \frac{1}{4}(\Delta s_1/s_1)$, with T_e in K). The amplitude of the SH contribution inferred from our analysis is lower (it never exceeds $\sim 3^\circ\text{C}$) but still compatible with our working hypothesis of a potential role of this local insolation. Note also that, even though never contributing $\geq 20\%$ in the ΔT reconstruction (Table 1), this local input plays a key role in bringing power in the obliquity band, which clearly dominates the Vostok temperature record¹. There is not, as might have been expected³⁰, a relation between the extension of sea-ice (if governed by the 60°S November insolation) and the temperature in central Antarctica.

Experiments with general circulation models (GCMs) show

that the Northern Hemisphere ice sheets have little influence on atmospheric and sea surface temperature (SST) in the Southern Hemisphere^{11,12}. Thus the atmosphere cannot constitute the suggested link between a NH input (65°N insolation or ice volume) and the Vostok climate. Such an interhemispheric link would instead be through a change in the deep oceanic circulation^{10,11,18–20}. A proposed mechanism¹⁹ is that under glacial conditions the formation of relatively warm North Atlantic deep water and its injection into circumpolar Antarctic regions are greatly reduced, thereby reducing the SH ocean temperature. Such changes in deep ocean circulation are now well documented all over the last climatic cycle^{38,44–48} and are in general agreement with this picture of a predominant role of the rate of formation of North Atlantic deep water. Sea-level variations associated with NH ice-sheet growth and retreat are also often cited as a potential link between NH and SH climates. We note that the suggestion that the global ice-sheet system interlocked by sea level (which transforms regional Milankovitch insolation into a global climate signal) is probably flawed with respect to the Southern Hemisphere²¹.

As noted above, no clear information about the phasing of the SH and NH climates, and thus about the potential link between the two hemispheres, can be extracted from our analysis. But it is of interest to examine whether such a specific climatic event as the Younger Dryas reversal thought to be typical of the NH climate⁴⁹ has a SH counterpart. As discussed in ref. 1, the amplitude of this cold reversal is much weaker in the Antarctic cores than in the Greenland cores, suggesting^{18,50} that it is a damped response of the SH climate to an initial Northern Atlantic cooling, hypothesized as cause of the Younger Dryas⁴⁹. If the present dating, which puts the Vostok cooling ~ 1 kyr earlier than the Younger Dryas (dated from 11 to 10.2 kyr BP), seems to contradict this suggestion, we are convinced that this event (unless related to CO₂ atmospheric content variation) is the manifestation of the NH–SH climatic link.

Our discussion of the contribution of CO₂ to climate is based on climate-sensitivity experiments made with GCMs either for modern^{33,34} or Last Glacial Maximum (LGM) boundary conditions^{11,12,51,52}. As noted above, such experiments allow us to study the various feedback processes influencing the climate sensitivity and to evaluate the net feedback factor, f , associated with a given change in radiative forcing. We examine whether the existence of such feedbacks may explain the high-CO₂ amplifications (~ 5 to ~ 14) found in this study.

The presentation of the GISS (Goddard Institute for Space Studies) model results^{33,34} is particularly well adapted to our purpose of examining feedback processes. With this model the predicted warming for $2\times\text{CO}_2$ experiments is $3\pm 1.5^\circ\text{C}$, which corresponds to a global-scale climate feedback of 2.4 ± 1.2 . A feature common to most of the GCM simulations is that they show greater changes in high latitudes than in low latitudes, linked to the variations in the amount of sea ice and extent of snow cover. This polar amplification was given as a possible reason for a significant impact in polar regions of the LGM decrease in atmospheric CO₂ content²⁶. But the temperature change predicted by the GISS model for modern $2\times\text{CO}_2$ experiments over central Antarctica is quite comparable with that predicted for the entire planet, showing no polar amplification in that region. Such a result does not support the idea that the high CO₂ feedback at Vostok inferred from our statistical analysis may be attributed only to polar amplification.

A more plausible explanation stems from the higher climate sensitivity during glacial than during interglacial periods, a point well documented by GCM models^{34,51,52}. The basic reason is that experiments with modern boundary conditions account only for fast feedback processes (water vapour, clouds and sea ice) and not for slow feedback processes such as those associated with the larger extent of sea ice and land ice during glacial periods. As noted by Hansen *et al.*³⁴, if strong positive feedbacks exist, a moderate additional feedback may cause a large increase

in the overall feedback factor. We can calculate the overall feedback from the property of additivity of the associated gains, $g = 1 - (1/f)$. From the results of Hansen *et al.* it follows that combining the land-ice feedback ($f = 1.25$) and that due to the additional sea ice ($f = 1.08$) with that resulting from the fast processes ($f = 2.4$) leads to an overall feedback of 7.

The same approach may be followed from the recent simulations of the LGM by Broccoli and Manabe¹² using the GFDL (Geophysical Fluid Dynamics Laboratory) model. Their experiments with a CO₂ decrease from 300 to 200 p.p.m.v. show a global temperature reduction of 1.2 °C but, in that case, the polar amplification affects all the Antarctic continent over which the temperature decrease is >2 °C (ref. 12). Indeed, the decrease in temperature due to the reduction of atmospheric CO₂ is responsible for almost all of the simulated LGM Southern Hemisphere cooling¹². The resulting feedback factor is >3 all over Antarctica (ΔT_{CO_2} is calculated as in ref. 34). In their experiments Manabe and Broccoli used LGM continental ice distribution so that the land-ice feedback was already included, but this may be counterbalanced by the fact that the effects of cloud feedback are not included. A moderate cloud feedback ($f = 1.15$, compared with 1.3 in the GISS model³⁴) and that due to additional sea ice (assumed to be 1.08) would also lead to a net feedback factor 7, at least over Antarctica.

These estimations do not include vegetation feedback (mentioned in ref. 34) and that postulated by Manabe and Bryan⁵¹ and Denton *et al.*²¹ involving respectively thermohaline circulation and floating shelf ice around Antarctica. It thus appears that, over long timescales, high CO₂ feedbacks as deduced from Vostok data are not unreasonable.

In that respect, on a global scale the direct radiative effect due to CO₂ changes (amplitude of ~0.6 °C) is about three times that due to insolation changes (the energy received by the planet varies by ~0.3% over the past 160 kyr, which would account for 0.2 °C). This supports our conclusion from multivariate analysis that CO₂ changes play a more important role than orbital variations. But the presence of the precession frequency in the CO₂ record² suggests that orbital forcing is also implicitly present in the CO₂ forcing.

Conclusion

The spectral characteristics of the CO₂ and climate records and the above discussion of climatic feedbacks lead to the idea that climatic changes would be triggered by insolation changes with the relatively weak orbital forcing being strongly amplified by

possibly orbitally induced CO₂ changes. This is in the general line of the ideas developed by Pias and Shackleton²⁷ in their modelling of the global climatic response to orbital forcing and CO₂ changes. Our interpretation must remain cautious as it does not shed light on the physical mechanisms involved; our results do not even allow us to exclude the possibility that CO₂ changes are just a consequence of climatic changes. However, the Vostok series appear to provide the most direct support so far of such an interaction between CO₂, orbital forcing and climate because (1) we have straightforward access to the climatic variables of interest (namely atmospheric CO₂ content and atmospheric temperature) and (2) a thorough examination of the climate model results does not contradict such a CO₂-amplifying role. Note also that such a climatic role for CO₂ would certainly make it easier to explain the synchronism of major glaciation in the Northern and Southern Hemispheres¹⁰⁻¹², one of the points left unanswered in the Milankovitch theory of the ice ages.

One can speculate from the Vostok results that the 100-kyr glacial-interglacial oscillation has its origin in the large associated CO₂ variations rather than in postulated nonlinearities associated with ice-sheet growth and decay^{8,9,13-17}. Indeed, a study of the non-glacial past clearly shows that the presence of ice sheets is not a necessary condition for the existence of the ~100-kyr cycle and that other interpretations are possible⁵³⁻⁵⁵.

Finally, we remain cautious about extrapolating from this analysis of past conditions to the impact of the recent anthropogenic increase in CO₂ on future climate. Even if our suggested interpretation is correct, one must realize that long-term feedbacks operate differently when going towards the last or the present interglacial than when going towards a possible future CO₂ superinterglacial. Such a prediction needs a better understanding of the mechanisms involved in this interaction between CO₂ orbital forcing and climate. We follow Broecker *et al.*¹⁸ in calling for the development and testing of joint ocean-atmosphere models as a tool to attack this complex but extremely important problem.

We thank A. Berger, W. S. Broecker, J. C. Duplessy, R. Fairbanks, J. E. Hansen, J. Imbrie, G. Kukla, L. D. Labeyrie, A. McIntyre, J. R. Petit, G. Russell and J. White for constructive criticisms. This work was supported in France by Expéditions Polaires Françaises, the Programme National d'Études de la Dynamique du Climat, Terres Australes et Antarctiques Françaises, and the Commission of the European Communities (Climatology Research Programme), and in the USSR by Soviet Antarctic Expeditions.

Received 26 March; accepted 17 July 1987.

- Jouzel, J. *et al.* *Nature* **329**, 408-414 (1987).
- Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 403-408 (1987).
- Lorius, C. *et al.* *Nature* **316**, 591-596 (1985).
- Milankovitch, M. K. *Acta Spec. Publ.* **133** (1941). (Translated by Israel Program for Scientific Translations, Jerusalem, 1969).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121-1132 (1976).
- Imbrie, J. *et al.* in *Milankovitch and Climate* Vol. 1 (eds Berger, A. L. *et al.*) 269-305 (Reidel, Dordrecht, 1984).
- Suarez, M. J. & Held, I. M. *J. geophys. Res.* **84**, 4825-4836 (1979).
- Imbrie, J. & Imbrie, J. Z. *Science* **207**, 243-253 (1980).
- Budd, W. F. *J.A.H.S. Publ.* **131**, 441-471 (1981).
- Broecker, W. S. in *Milankovitch and Climate* Vol. 2 (eds Berger, A. L. *et al.*) 687-698 (Reidel, Dordrecht, 1984).
- Manabe, S. & Broccoli, A. J. *J. geophys. Res.* **90**, 2167-2190 (1985).
- Broccoli, A. J. & Manabe, S. *Clim. Dynam.* **1**, 87-99 (1987).
- Birchfield, G. E., Weertman, J. & Lunde, A. T. *Quat. Res.* **15**, 126-142 (1981).
- Pollard, D. *Nature* **296**, 334-338 (1981).
- Budd, W. F. & Smith, I. N. *J.A.H.S. Publ.* **131**, 369-409 (1981).
- Le Treut, H. & Ghil, M. *J. geophys. Res.* **88**, 5167-5190 (1983).
- Lindzen, R. S. *J. Atmos. Sci.* **43**, 986-996 (1986).
- Broecker, W., Peteet, D. & Rind, D. *Nature* **315**, 21-26 (1985).
- Corliss, B. H. *Nature* **298**, 458-460 (1982).
- Johnson, R. G. & Andrews, J. T. *Paleogeogr., Palaeoclimatol., Palaeoecol.* **5**, 107-138 (1986).
- Denton, G. H., Huges, T. J. & Karlen, W. *Quat. Res.* **26**, 3-26 (1986).
- Delmas, R. J., Ascencio, J. M. & Legrand, M. *Nature* **284**, 155-157 (1980).
- Nefel, A., Oeschger, H., Schwander, R. J., Stauffer, B. & Zumbund, R. *Nature* **295**, 220-223 (1982).
- Shackleton, N. J., Hall, M. A., Line, J. & Cag Shuzi. *Nature* **306**, 319-323 (1983).
- Shackleton, N. J. & Pias, N. G. *Geophys. Monogr.* **32** (eds Sundquist, E. T. & Broecker, W. S.) 303-317 (American Geophysical Union, Washington DC, 1985).
- Stauffer, B., Hofer, H., Oeschger, H., Schwander, J. & Siegenthaler, U. *Ann. Glaciol.* **5**, 160-164 (1984).
- Pias, N. G. & Shackleton, N. J. *Nature* **30**, 757-759 (1984).
- Moriyama, S. *J. met. Soc. Japan* **64**, 613-621 (1986).
- Berger, A. L. *J. Atmos. Sci.* **35**, 2362-2367 (1978).
- Kukla, G. in *Climatic Change* (ed. Gribbin, J.) 114-129 (Cambridge University Press, 1978).
- Carroll, J. J. & Fitch, W. B. *J. geophys. Res.* **86**, 5271-5276 (1981).
- Martinson, D. G. *et al.* *Quat. Res.* **27**, 1-29 (1987).
- Hansen, J. *et al.* *Science* **220**, 857-859 (1985).
- Hansen, J. *et al.* *Geophys. Monogr.* **20** (M. Ewing symposium 5) 130-163 (American Geophysical Union, Washington DC, 1984).
- Siegenthaler, U. & Oeschger, H. *Ann. Glaciol.* **5**, 153-159 (1984).
- Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Application* (Holden-Day, San Francisco, 1968).
- Imbrie, J. *J. geol. Soc. Lond.* **142**, 417-432 (1985).
- Labeyrie, L., Duplessy, J. C. & Blanc, P. L. *Nature* **327**, 477-482 (1987).
- Rasool, S. I. & Schneider, S. H. *Science* **173**, 138-141 (1971).
- Wang, W. C., Wuebbles, D. J., Washington, W. M., Isaac, R. G. & Molinar, G. *Rev. Geophys.* **24**, 110-140 (1986).
- De Angelis, M., Barkov, N. I. & Petrov, V. N. *Nature* **325**, 318-321 (1987).
- Barrodale, I. & Erickson, R. E. *Geophysics* **45**, 420-432 (1980).
- Ghil, M. in *Climatic Variations and Variability: Fact and Theory* (ed. Berger, A.) 461-480 (Reidel, Dordrecht, 1981).
- Duplessy, J. C., Moyes, J. & Pujol, C. *Nature* **286**, 479-482 (1980).
- Shackleton, N. J., Imbrie, J. & Hall, M. A. *Earth planet. Sci. Lett.* **65**, 233-244 (1983).
- Mix, A. C. & Fairbanks, R. G. *Earth planet. Sci. Lett.* **73**, 231-243 (1985).
- Duplessy, J. C. & Shackleton, N. J. *Nature* **316**, 500-507 (1985).
- Oppo, D. W., Fairbanks, R. G. & Mix, A. *Quat. Res.* (submitted).
- Rind, D., Peteet, D., Broecker, W., McIntyre, A. & Ruddiman, W. *Climate Dynam.* **1**, 3-33 (1986).
- Jouzel, J., Lorius, C., Merlivat, L. & Petit, J. R. *Abrupt Climatic Changes—Evidences and Implications*, 235-245 (Reidel, Dordrecht, 1987).
- Manabe, S. & Bryan, K. *J. geophys. Res.* **90**, 11689-11707 (1985).
- Manabe, S. & Broccoli, A. J. *J. Atmos. Sci.* **42**, 2643-2651 (1985).
- Herbert, T. D. & Fisher, A. G. *Nature* **321**, 739-743 (1986).
- Olsen, P. E. *Science* **23**, 842-848 (1986).
- Mix, A. *Nature* **327**, 370 (1987).

The polarization spectrum of supernova 1987A interpreted in terms of shape asymmetry

D. J. Jeffery

Physics Department, McMaster University, Hamilton, Ontario L8S 4M1, Canada

Polarimetry done by Schwarz and Mundt¹ on the type II supernova 1987A on 6 and 7 March 1987 showed variation in polarization across line profiles (see Table 1). This polarization structure is interpreted as arising from an asymmetric, homologously expanding, scattering atmosphere surrounding an asymmetric continuum-producing photosphere. Resonant scattering of radiation by ions in the atmosphere produces the line structure in the flux spectrum and polarizes the emergent radiation. The asymmetric shape of the atmosphere causes a non-zero net polarization. Sobolev-method radiative transfer calculations with axisymmetric oblate ellipsoidal models have been carried out to fit the observed data. The models are parameterized by the ratio of the symmetry axis to the perpendicular axis, c/a . The fits to the 1987A data indicate that (c/a) is $\sim 0.6-0.8$.

The Sobolev method²⁻⁴ is used for calculating resonant line spectra in atmospheres with large velocity gradients and has been successfully applied to supernovae by Branch *et al.*⁵⁻⁷. The usual Sobolev method relies on complete redistribution of photons in frequency and angle. Complete redistribution (CRD)

in frequency means that the incoming and outgoing photon frequencies are uncorrelated and are described by independent absorption and emission profiles. CRD in angle means that the scattering is isotropic and non-polarizing. To include polarization a hybrid redistribution was considered. This hybrid redistribution has complete redistribution in frequency, but to give the angular redistribution of the scattered radiation it includes a linear combination of the isotropic and Rayleigh phase matrices (see Chandrasekhar⁸ for the prescription of the matrix coefficients). The radiation fields are described in terms of the Stokes parameters. The following expressions were obtained for the Stokes parameter source functions assuming axial symmetry and no circular polarization:

$$S_l(r) = S_1 + S_2\mu^2 + S_3\mu\sqrt{1-\mu^2}\cos\phi + S_4\mu^2\cos 2\phi \quad (1a)$$

$$S_r(r) = S_5 - S_4\cos 2\phi \quad (1b)$$

$$S_u(r) = S_3\sqrt{1-\mu^2}\sin\phi + 2S_4\sin 2\phi \quad (1c)$$

where l indicates the radiation vector aligned with the axis of symmetry, r the vector aligned perpendicular to the axis of symmetry, and u the difference between the l and r components in a 45° rotated system. The coefficients labelled 1-5 are obtained by numerical integration over the incoming specific intensity, and by some linear algebra.

The models examined with this modified Sobolev method consist of an axially symmetric ellipsoid photosphere which produces an unpolarized, frequency-independent or Planck continuum, and a similar ellipsoid atmosphere in which there is only resonant scattering by ions. The Sobolev optical depths

Table 1 Polarimetric data for SN1987A

Feature	λ (Å) [†]	$Q(\%)$	$U(\%)$	$\chi(^{\circ})$	$P(\%)$
ISP‡		$+0.88 \pm 0.07$	$+0.41 \pm 0.07$	12.5 ± 2.7	0.97 ± 0.09
GFP		$+0.24 \pm 0.07$	$+0.18 \pm 0.07$	18.4 ± 9.3	0.30 ± 0.10
UBVRI (25 February 1987)§		-0.02 ± 0.07	-0.14 ± 0.07	131 ± 16	0.14 ± 0.08
UBVRI (28 February 1987)		-0.10 ± 0.07	-0.04 ± 0.07	101 ± 24	0.11 ± 0.09
UBVRI (2 March 1987)		-0.19 ± 0.07	-0.09 ± 0.07	103 ± 13	0.21 ± 0.09
UBVRI (7 March 1987)		-0.36 ± 0.07	-0.09 ± 0.07	97 ± 7	0.37 ± 0.08
H α (peak)	6,565 10	-0.12 ± 0.07	-0.08 ± 0.07	107 ± 19	0.14 ± 0.10
H α (trough)	6,251 33	-0.36 ± 0.07	-0.24 ± 0.07	107 ± 6	0.43 ± 0.10
H β (peak)	4,867 34	-0.32 ± 0.07	-0.21 ± 0.07	107 ± 7	0.38 ± 0.10
H β (trough)	4,697 10	-0.94 ± 0.07	-0.69 ± 0.07	108 ± 2	1.17 ± 0.10
H γ (peak)	4,340 28	-0.30 ± 0.07	-0.02 ± 0.07	92 ± 7	0.30 ± 0.07
H γ (trough)	4,188 33	-0.46 ± 0.07	-0.41 ± 0.07	111 ± 5	0.62 ± 0.10
Na D (peak)	5,897 56	-0.26 ± 0.07	-0.07 ± 0.07	98 ± 9	0.27 ± 0.09
Na D (trough)	5,757 20	-0.60 ± 0.07	-0.44 ± 0.07	108 ± 4	0.74 ± 0.10

Unless otherwise indicated the data were taken on the nights of 6 and 7 March 1987 (ref. 1).

[†] These values describe the filters used by Schwarz and Mundt¹: for example, central wavelength/full width half maximum.

[‡] ISP is the interstellar polarization and GFP is the galactic foreground polarization; ISP value from Barrett¹⁰; the other features have been corrected by these values to give intrinsic polarizations.

[§] UBVRI labels the mean polarization of the broadband colours U, B, V, R and I.

Table 2 Spectroscopic line data for SN1987A taken on 6 March 1987

Line	λ_{peak} (Å)	F_{peak}^*	λ_{trough} (Å)	F_{trough}	$\Delta\lambda$ (Å)	V_{trough} (cm s^{-1} $\times 10^3$)	$\gamma_{\mu}f^{\dagger}$	τ_{ph}
H α	6,562	1.95 ± 0.10	6,284	0.25 ± 0.10	278	1.27	0.4205	1,000
H β	4,861	1.10 ± 0.10	4,697	0.20 ± 0.10	170	1.05	0.05799	138
H γ	4,340	0.55 ± 0.10	4,188	0.20 ± 0.10	152	1.05	0.01938	46
Na D	5,889	1.00 ± 0.10	5,757	0.60 ± 0.10	140	0.712	0.577	3.1

Data from ref. 9.

* The flux F is normalized to the continuum value at 5,900 Å.

[†] The Sobolev optical depths for a supernova, neglecting stimulated emission, are given by $\tau = 0.23\lambda_{\mu}fNt_d$, where λ_{μ} is the wavelength in micrometres, f is the oscillator strength, N is the density of absorbers in the lower level of the transition per cubic centimetre and t_d is the time in days since the explosion. Thus the optical depths for the Balmer lines always have the same ratio in this approximation.

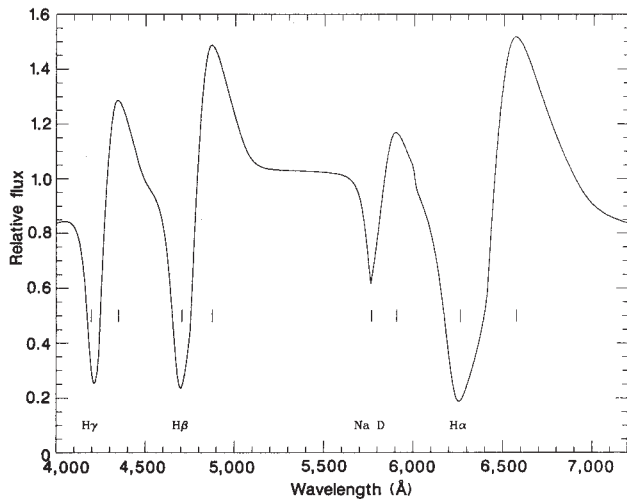


Fig. 1 A spherically symmetric model synthetic flux spectrum for 1987A on 6 March 1987 showing H α , H β , H γ and Na D lines. The fitted τ_{ph} values are 1,000, 138, 46 and 3.1 respectively. The line-of-sight velocity of that part of the photosphere nearest the observer is set to $0.71 \times 10^9 \text{ cm s}^{-1}$. This value was obtained by fitting the wavelength shift of the weak Na D line's absorption trough⁹. The background continuum is a Planck spectrum with $T = 5,500 \text{ K}$. This value was obtained by interpolating 1 March and 11 March Planck spectrum fits¹⁵. The flux is normalized to its continuum value at 5,900 Å. The fitted τ_{ph} values from this model are used for the asymmetric models because the flux profiles are not changed drastically by varying (c/a) for $(c/a) \geq 0.5$.

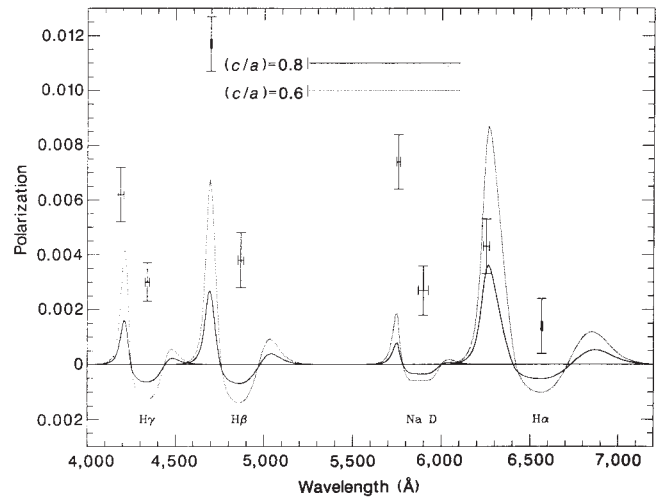


Fig. 2 Synthetic line polarization spectra for supernova 1987A (6 and 7 March 1987) for oblate ellipsoid models. The inclination angle of the symmetry axis was set to 90° (see Fig. 1 legend for a description of the other model parameters). Only the $(c/a) = 0.8$ and $(c/a) = 0.6$ model spectra are shown for clarity. The Schwarz and Mundt¹ data points are shown without any correction for a continuum polarization. Vertical error bars, measurement uncertainty; horizontal error bars, half-width of the filters.

were parameterized according to

$$\tau = \tau_{ph} \left(\frac{g_{ph}}{g} \right)^p \quad (4)$$

τ_{ph} , a fitting parameter, is the optical depth at the photosphere, and p , in correspondence to the choice of Branch *et al.* for spherically symmetric models⁵⁻⁷, is set to 7. The factor g is a generalized radius parameter given by

$$g = \sqrt{(x/a)^2 + (y/a)^2 + (z/c)^2} \quad (5)$$

where c and a are, respectively, the semiaxes parallel and perpendicular to the axis of symmetry.

A parameter survey of asymmetric models allows several general conclusions to be drawn that apply to 1987A. (1) There is polarization structure associated with the P-Cygni line profiles. (2) There is in most cases an inversion in the polarization profile (a 90° shift in polarization angle) between the P-Cygni peak and trough features for both oblate and prolate models. The peak polarization vector aligns parallel to the long axis of the model and the trough polarization vector is perpendicular to this axis. (3) Prolate and oblate models for the same degree of asymmetry, and with $2 \geq (c/a) \geq 0.5$, have similar degrees of polarization. (4) The magnitude of trough polarization is typically more than twice that of peak polarization. (5) The trough feature is typically narrower than the peak feature. (6) The

polarization values obtained are not very large even for large asymmetries. For example, for a typical prolate ellipsoid model the difference between peak and trough polarization is $\sim 1.7\%$ when $(c/a) = 2$ and with other parameters chosen to maximize polarization. For a typical oblate ellipsoid model with $(c/a) = 0.5$ the difference between peak and trough polarization is also $\sim 1.7\%$.

There is spectropolarimetric data for four of 1987A's lines¹: H α , H β , H γ and Na D (see Tables 1 and 2). None of the lines was a pure, unblended resonant scattering line on 6 March⁹. The H α line shows more emission than absorption, while the reverse is true of a resonant scattering line. There may be a thermal source of photons and this would tend to depolarize this line's flux. The H β , H γ and Na D lines are not unblended, although the absorption features of the H β and Na D line are well defined. A synthetic spectrum for these lines for a spherically symmetric model is given in Fig. 1. The fitted τ_{ph} values from this model have been used for the asymmetric model calculations because the shapes of the flux profiles do not change drastically for values of $(c/a) \geq 0.5$. The inclination angle of the symmetry axis of the asymmetric models was set to 90° because the large polarization of the H β line could not be fitted by substantially smaller angles. Figure 2 shows polarization profiles for oblate models with varied (c/a) . Oblate models were studied because oblateness has an obvious physical cause in the rotation of a supernova progenitor.

The polarimetric data¹ (corrected for interstellar polarization by using Barrett's value¹⁰) do not exhibit the polarization inversion predicted by the Sobolev models (see Table 1). It may be

Table 3 Observational and model ΔP values

Line	$\Delta P_{obs} (\%)$	$\Delta P_{model} (\%)*$				
		$(c/a) = 1$	$(c/a) = 0.8$	$(c/a) = 0.6$	$(c/a) = 0.5$	$(c/a) = 0.25$
H α	0.29 ± 0.20	0.00	0.40 ± 0.03	0.90 ± 0.03	1.17 ± 0.03	1.61 ± 0.03
H β	0.79 ± 0.20	0.00	0.33 ± 0.02	0.79 ± 0.02	1.09 ± 0.02	2.00 ± 0.02
H γ	0.32 ± 0.17	0.00	0.16 ± 0.04	0.36 ± 0.04	0.46 ± 0.04	0.56 ± 0.04
Na D	0.47 ± 0.19	0.00	0.10 ± 0.03	0.21 ± 0.03	0.26 ± 0.03	0.32 ± 0.03

* The uncertainties assigned to the ΔP_{model} values are only the numerical uncertainties.

that there is a continuum polarization upon which the line polarization profiles are superimposed. Assuming this to be the case the peak-to-trough polarization difference ΔP becomes the quantity to be fitted. The ΔP values for the observational and model lines are given in Table 3. Recalling that none of the lines is a pure resonant scattering line, it is not surprising that the best fitting (c/a) values are different for each line. But the Balmer lines are fitted by (c/a) values in the range 0.6–0.8 and the conclusion is drawn that supernova 1987A is 20 to 40% oblate asymmetric.

Shapiro and Sutherland¹¹ considered continuum polarization from ellipsoidal supernovae and McCall¹², line polarization. Their physical models are somewhat different from that of this letter. But from their results, estimates of 1987A's asymmetry can be made that are also ~20–40%.

Müller and Hillebrandt's¹³ rotating core explosion models show core oblateness of $\geq 20\%$ for core rotational energies $\geq 10^{49}$ erg. Rotational energies of a few 10^{49} erg are attributed to pulsars¹⁴, the presumed remnants of type-II supernovae. The core oblateness may be communicated to the atmosphere, where it would become the source of the polarization. The details of Sobolev polarization calculations and further fits will appear in a paper now in preparation.

I thank Ralph Pudritz and Peter Sutherland for advice and encouragement, and David Branch, for suggestions.

Received 8 June; accepted 15 July 1987.

1. Schwarz, H. E. & Mundt, R. *Astr. Astrophys.* **177**, L4–L4 (1987).
2. Sobolev, V. V. *Moving Envelopes of Stars* (Harvard University Press, Cambridge, 1960).
3. Rybicki, G. B. & Hummer, D. G. *Astrophys. J.* **219**, 654–675 (1978).
4. Olson, G. L. *Astrophys. J.* **255**, 267–277 (1982).
5. Branch, D. in *Supernovae Spectra* (eds Meyerott, R. & Gillespie, G. H.) 39–48 (American Institute of Physics, New York, 1980).
6. Branch, D. *et al. Astrophys. J.* **252**, L61–L64 (1982).
7. Branch, D., Dogget, J. B., Nomoto, K. & Thielemann, F.-K. *Astrophys. J.* **294**, 619–625 (1985).
8. Chandrasekhar, S. *Radiative Transfer* (Dover, New York, 1960).
9. Danziger, I. J. *et al. Astr. Astrophys.* **177**, L13–L16 (1987).
10. Barrett, P. *IAU Circ.* no. 4337 (1987).
11. Shapiro, P. R. & Sutherland, P. G. *Astrophys. J.* **263**, 902–924 (1982).
12. McCall, M. L. in *Supernovae as Distance Indicators* (ed. Bartel, N.) 48–61 (Springer, Berlin, 1985).
13. Müller, E. & Hillebrandt, W. *Astr. Astrophys.* **103**, 358–366 (1981).
14. Gunn, J. E. & Ostriker, J. P. *Nature* **21**, 454–456 (1969).
15. Bouchet, R. *et al. Astr. Astrophys.* **177**, L9–L12 (1987).

Modelling of the radio burst from SN1987A

Michelle C. Storey & R. N. Manchester

Division of Radiophysics, CSIRO, PO Box 76, Epping, NSW 2121, Australia

Observations at Australian observatories¹ showed that radio emission from SN1987A in the Large Magellanic Cloud² reached a peak of ~100 mJy at frequencies of ~1 GHz within three days of the optical rise and then declined on a similar timescale. This behaviour is different from that observed from previous extragalactic type II radio supernovae³ where the peak luminosity is two to three orders of magnitude greater and the emission rises tens or even hundreds of days after the supernova event and then decays with a timescale of years. Here we propose that the radio emission mechanism for SN1987A is the same as that proposed for previous type II supernova⁴, namely, synchrotron emission from a thin shell near the edge of the expanding ejecta. The differences are mainly due to the absence of a thick circumstellar envelope around the progenitor star, Sk -69 202, at least in the line of sight to the supernova. We discuss several possible models for the emission and absorption processes and set limits on the density of a circumstellar envelope.

We assume that the relativistic electrons were located in a shell of radius R and thickness D located at the outer edge of

the expanding ejecta. The shell is assumed to expand with $R \propto t^m$ where t is the time since the supernova explosion and m is a constant < 1 allowing for deceleration of the shell expansion (see ref. 4). The relativistic electrons could have been accelerated during the explosion by, for example, first-order Fermi acceleration as the supernova shock propagated through the stellar atmosphere (see ref. 5). Magnetic fields would have been trapped in the expanding ejecta and on day 2 (25 February) field strengths in the shell of the order of a milligauss are reasonable. For synchrotron radiation in the radio band, electron Lorentz factors $\gamma_e \sim 100$ are implied and the gyro-radius is much less than the thickness of the shell, so the relativistic electrons would be contained by the magnetic field in the shell. The relativistic electrons are assumed to have a power-law differential energy distribution

$$N(E) dE = KE^{-\gamma} dE \quad (1)$$

The synchrotron radiation then has an emissivity of

$$j_\nu \propto a(\gamma) KB^{(\gamma+1)/2} \nu^{-(\gamma-1)/2} \quad (2)$$

where $a(\gamma)$ is a slow function of γ , B is the magnetic flux density and ν is the radio frequency⁶. To produce the observed radio luminosity of $\sim 2 \times 10^{33}$ erg s⁻¹, the total energy in the synchrotron electrons is $\sim 10^{44}$ erg, a minute fraction of the total energy output of the supernova.

The relativistic electrons will suffer energy losses due to a number of processes. Synchrotron losses are unimportant having a timescale of thousands of years. Adiabatic losses have a timescale of RV^{-1} , where V is the expansion velocity, and hence are significant on the timescale of the event. In the intense radiation field of the supernova, inverse Compton losses are likely to be important. The timescale for this process τ_{IC} scales as $R^2(\gamma_e L_r)^{-1}$, where L_r is the radiative luminosity of the supernova. Initially L_r was dominated by optical and ultraviolet emission and was $\sim 10^{41}$ erg s⁻¹. For $R \sim 10^{14}$ cm (appropriate for day 2) and $\gamma_e \sim 100$, $\tau_{IC} \sim 3 \times 10^5$ s. Therefore, at times of interest, adiabatic losses dominate. It is possible that the energy losses due to these processes continued to be overcome by generation of relativistic electrons (see ref. 4) and we consider models where this is and is not the case.

Absorption processes which may be important include free-free absorption and synchrotron self-absorption. The successful interpretation of previous extragalactic type II supernovae^{3,4} is based on free-free absorption in a circumstellar envelope with an r^{-2} density profile, presumably generated during a red giant phase of the progenitor star. For SN1987A any circumstellar envelope must be relatively thin, at least in the line of sight to the supernova, otherwise the prompt radio burst would not have been observed. The progenitor star for SN1987A, Sk -69 202, was apparently a B3 supergiant star⁷ which would not be expected to have an extensive circumstellar envelope, and indeed, there is no evidence for one in either the optical or ultraviolet spectra^{8,9}. The coefficient of free-free absorption can be approximated by

$$k_{ff} = 0.108 T_e^{-3/2} n_{th}^2 \nu^{-2} \text{ cm}^{-1} \quad (3)$$

where T_e is the electron temperature and n_{th} is the thermal electron density¹⁰. For T_e of a few thousand degrees, a limit on $n_{th} \leq 10^6$ cm⁻³ in or near the emission region on day 2 is implied by the requirement that the free-free optical depth be ≤ 1 at frequencies of ~1 GHz. The emission region must therefore be close to the outer edge of the expanding ejecta and any circumstellar envelope must be thin.

If thermal plasma exists in the emission region, Razin-Tsytovich suppression¹⁰ of the synchrotron emission may be important. The criterion for this not to be the case is that $\nu > \gamma_e \nu_p$ where ν_p is the plasma frequency. For day 2, this implies a limit on n_{th} similar to that obtained for free-free absorption. Therefore we do not expect this effect to be important at the times of interest. Finally, synchrotron self-absorption may be significant.

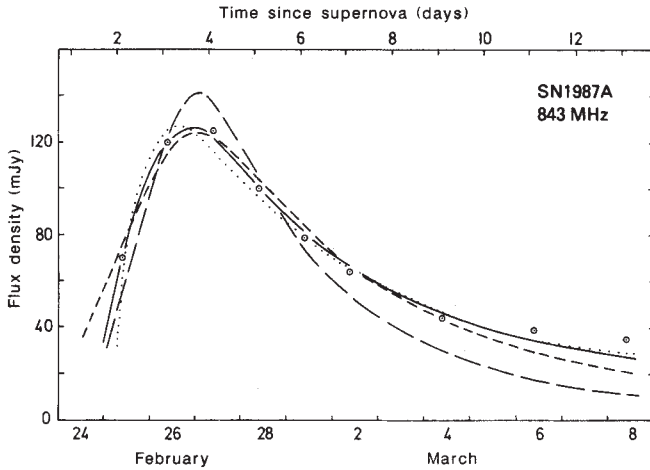


Fig. 1 Observed time dependence of the 843-MHz flux density of SN1987A from Turtle *et al.*¹ and the fits of four different models to these data. Observational errors are ± 2 mJy. For models 1 and 2, both thermal and relativistic electrons are contained in a relatively thin expanding shell of constant thickness located near the edge of the ejecta. Free-free absorption and synchrotron self-absorption are included in the fits. For model 1, the coefficient K is assumed to vary as R^{-2} , whereas in model 2 adiabatic expansion of the relativistic electrons is assumed. Model 3 is similar to model 2 except that the emitting and absorbing shell is assumed to expand with $D \propto R$. For model 4, free-free absorption occurs in a r^{-2} circumstellar envelope and all other absorption processes are neglected. The assumptions for the synchrotron-emitting shell are as in model 1. —, Model 1; ---, model 2; —·—, model 3; ·····, model 4.

The absorption coefficient for this process is given by⁶

$$k_{\text{syn}} \propto g(\gamma) K B^{(\gamma+2)/2} \nu^{-(\gamma+4)/2} \quad (4)$$

where $g(\gamma)$ is a slowly varying function of γ .

We consider four models for the prompt radio burst. For each model, we have made a least-squares fit to the 843-MHz data from the Molonglo Observatory¹. These data have the highest sensitivity of those available and show both the rise and the fall of the emission most clearly. Time-dependent parameters for the models are quoted at time t_0 , 25.4 February UT or 2.1 days after the supernova explosion, the time of the first Molonglo observation. For all fits, the deceleration parameter $m = 0.99$, corresponding to essentially free expansion.

In models 1 and 2, thermal electrons are assumed to coexist with the synchrotron electrons in an expanding shell of constant thickness $D = \zeta R_0$ where $R_0 = R$ at t_0 . Any external wind envelope is assumed to be of negligible density. The magnetic field is assumed to be random and hence $B \propto R^{-1}$. It is further assumed that, as the shock went through the outer layers of the star, the shocked gas was able to cool and that subsequently the expansion was adiabatic and hence $T_e \propto R^{-4/3}$ for the times of interest. At t_0 a value of 5,000 K is assumed. The optical depth for free-free absorption along a radial path is then

$$\tau_{\text{ff}} = k_{\text{ff}} D \propto \zeta R^{-2} \nu^{-2} \quad (5)$$

where the constant of proportionality is determined by R , n_{th} and T_e at the reference time t_0 .

For model 1, the coefficient of the relativistic electron energy distribution, K , is assumed to vary as R^{-2} . This requires that acceleration processes continue to operate as the shell expands and have a short timescale compared to the dissipative processes (see ref. 4). The index γ is assumed to remain constant during the expansion. The optical depth for synchrotron self-absorption is then

$$\tau_{\text{syn}} = k_{\text{syn}} D \propto \zeta g(\gamma) K_0 B_0^{(\gamma+2)/2} R^{-(\gamma+6)/2} \nu^{-(\gamma+4)/2} \quad (6)$$

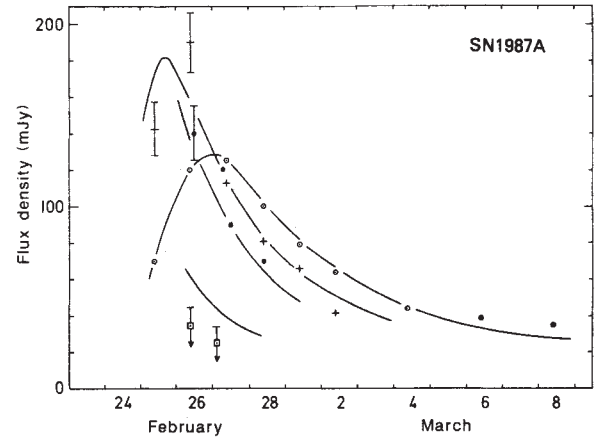


Fig. 2 Predicted flux densities from model 1 compared to the observations at all four radio frequencies. The apparent time of the supernova event^{12,13} was 23.3 February. Observations at 843 MHz ($\circ$), 2.3 GHz ($\bullet$) and 8.4 GHz ($\square$) are from Turtle *et al.*¹. Flux densities at 1.4 GHz (+) from the Fleurs observations have been revised upward by $\sim 35\%$ (J. D. Bunton, personal communication). Representative error bars are given for some of the points.

where K_0 and B_0 are the values of K and B at t_0 , respectively. From the equation of radiative transfer, the observed flux density is

$$S = \pi (R/d)^2 (j_\nu/k_\nu) (1 - e^{-\tau}) \quad (7)$$

(neglecting geometric effects), where d is the distance to the supernova, assumed to be 55 kpc, $k_\nu = k_{\text{syn}} + k_{\text{ff}}$ and $\tau = \tau_{\text{syn}} + \tau_{\text{ff}}$. The observed flux density rises with time until $\tau \sim 1$ due to the dominant effect of the decrease in absorption and then falls due to the decreasing synchrotron emissivity as the emitting shell expands.

For model 2, we assume that there is no further acceleration of the relativistic electrons and that their expansion is adiabatic at the times of interest. For this case γ remains constant during the expansion and $K \propto R^{-(2\gamma+4)/3}$. Then $j_\nu \propto R^{-(7\gamma+11)/6}$ and $k_{\text{syn}} \propto R^{-7(\gamma+2)/6}$.

For an essentially free expansion, one might expect the synchrotron-emitting shell to expand with thickness proportional to radius. Model 3 is the same as model 2 except that $D \propto R$. For this case, $K \propto R^{-(\gamma+2)}$, $B \propto R^{-2}$, $n_{\text{th}} \propto R^{-3}$ and $T_e \propto R^{-2}$. Hence $j_\nu \propto R^{-(2\gamma+3)}$, $\tau_{\text{ff}} \propto R^{-2}$ and $\tau_{\text{syn}} \propto R^{-(2\gamma+3)}$.

Finally, we take the model of Chevalier⁴ in which free-free absorption occurs in a circumstellar wind with $n_e \propto r^{-2}$. We assume $T_e = 5,000$ K in the wind. Absorption due to other processes is assumed negligible and K is taken to be $\propto R^{-2}$ as in model 1. As the shock expands the free-free optical depth in the line of sight from the shock front decreases as R^{-3} .

Figure 1 shows the fits of these models to the Molonglo data and the derived parameters are given in Table 1. In all four models the rapid rise and more gradual fall of the observed emission is reproduced. Model 1 gives the best fit with only the last two points being significantly off the fitted curve. Optical depths are small by day 4, so after this time the spectral index will be $\alpha = -(\gamma-1)/2 \sim -0.7$ in accordance with the observations¹. Model 2, where the relativistic electrons are assumed to expand adiabatically, also gives an acceptable fit with the thickness of the emitting shell being about three times larger than for model 1. For model 3 it was necessary to hold ζ constant to obtain convergence of the least-squares procedure. This model, where $D \propto R$, is clearly inferior with both the rise and fall of the emission being too steep. This more rapid evolution results from the stronger R -dependence of the various parameters. Both the relativistic and thermal electron densities are much higher in this model and the absorption optical depths

Table 1 Parameters from least-squares fits to Molonglo data

	V_0 (km s ⁻¹)	ζ	K_0 (c.g.s. units)	γ	B_0 (gauss)	n_{th0} (10 ⁵ cm ⁻³)
Model 1	18,600	0.048	8.7	2.33	0.0028	5.32
Model 2	22,100	0.173	16.2	2.06	0.0045	4.52
Model 3	21,200	0.200*	46.1	2.23	0.0138	16.80
Model 4	21,900	0.249	24.9	1.79	0.0056	1.73

* Fixed parameter.

are large for a long time and so this model does not give a good fit to the observed spectrum. For model 4, the external r^{-2} wind case, we had to set arbitrarily the synchrotron self-absorption to zero, as the least-squares fit gave a negligible density for the external wind if this was not done. This indicates that the model is not a good fit to the data as can be seen from Fig. 1. The exponential rise is too steep and it is not possible to fit the first three data points within their errors. Because the model 4 assumptions about the emission zone are similar to those for model 1, this model gives a good fit at later times when the free-free optical depth is small.

In Fig. 2 we show the flux densities predicted at the other observed frequencies using model 1 and the corresponding parameters given in Table 1. Overall, the model gives a good prediction of the observed flux densities (given that the error bars at the other frequencies are much larger than those at 843 MHz), especially at late times when absorption processes are negligible. The main discrepancies are that the predicted peak at 1.4 GHz occurs ~ 12 h too early and that the predicted flux densities at 8.4 GHz are $\sim 80\%$ too high. The observed later peak at 1.4 GHz implies a larger effective absorption coefficient at higher frequencies, at least initially, perhaps as a result of an initial relaxation of the density profiles in the emitting region. The lower flux densities at 8.4 GHz imply a steepening of the relativistic electron energy distribution at high energies which may result directly from the acceleration mechanism, but probably reflects the effects of energy-loss processes such as inverse Compton scattering. In addition, in all models, the derived expansion velocity $V_0 \sim 20,000$ km s⁻¹ is less than the maximum blueshift observed in optical⁸ and ultraviolet⁹ spectra. Further detailed modelling including more realistic density profiles for the ionized and relativistic gas is required to assess the significance of these discrepancies. The derived magnetic flux densities of a few milligauss at t_0 imply a stellar field of the order of a gauss, not inconsistent with that expected for a B3 star¹¹. The overall agreement of the model with the observations illustrated in Fig. 2 indicates that the basic model, synchrotron emission from an expanding shell together with free-free absorption from plasma in the emission region and/or synchrotron self-absorption, is correct.

Perhaps the most significant result from this study is that the density of thermal electrons, either in the emission region or in the line of sight to the supernova, is small. In particular, the value of n_{th0} quoted in Table 1 for model 4, the circumstellar wind model, is an upper limit on the actual density because other absorption processes have been neglected. If we assume the wind is fully ionized, we can take this as a limit on the gas density. The implied limit on mass-loss rate for the progenitor star is proportional to the assumed wind velocity; for 100 km s⁻¹ it is $10^{-7} M_\odot$ yr⁻¹, much smaller than the mass-loss rate expected from a red giant. This indicates either that the progenitor star did not go through a red giant phase or, if it did, that the mass outflow was nonspherical and such that our line of sight does not intercept the bulk of the circumstellar material.

This apparent lack of an extensive circumstellar envelope also accounts for the low luminosity of the prompt radio burst from SN1987A. In type II supernovae where there is a massive circumstellar envelope, the synchrotron electrons are probably generated by turbulent interactions as the shock encounters the envelope. For SN1987A the low circumstellar density may have

resulted in a very different shock. If a massive but non-spherical circumstellar envelope exists, the shock may still encounter this, perhaps resulting in a radio burst similar to those observed in previous type II SN. As mentioned by Turtle *et al.*¹, the prompt burst observed for SN1987A could not have been detected for any previous event.

There may be other, more centrally located, sources of radio radiating electrons. Relativistic electrons generated deeper within the ejecta by the initial passage of the shock and still behind absorbing layers lose energy through Coulomb and Compton collisions and no longer radiate at radio frequencies. However, a central energy source such as a pulsar may continue to generate relativistic electrons and the synchrotron radiation from such electrons may become observable at a later date. If there is no such central energy source and no massive circumstellar envelope it is possible that there will be no significant radio emission until (portions of) the ejecta with sufficient energy reach regions of the interstellar medium which are sufficiently dense for turbulent acceleration of electrons to occur. If the supernova is surrounded by a large cavity, this may never happen. We then have the interesting possibility of a supernova which apparently formed a neutron star^{12,13} having no associated radio remnant.

We thank J. A. Roberts and F. D. Kahn for helpful discussions. M.C.S. acknowledges the support of an Australian Government National Research Fellowship and R.N.M. acknowledges the hospitality of the Nuffield Radio Astronomy Laboratories and the support of an SERC Visiting Fellowship for a portion of this work.

Received 21 May; accepted 10 August 1987.

1. Turtle, A. J. *et al. Nature* **327**, 38–40 (1987).
2. Shelton, I. *IAU Circ. No.* 4316 (1987).
3. Weiler, K. W., Sramek, R. A., Panagia, N., van der Hulst, J. M. & Salvati, M. *Astrophys. J.* **301**, 790–812 (1986).
4. Chevalier, R. A. *Astrophys. J.* **259**, 302–310 (1982).
5. Blandford, R. D. & Eichler, D. *Phys. Rep.* (in the press).
6. Ginzburg, V. L. & Syrovatskii, S. I. *A. Rev. Astr. Astrophys.* **3**, 297–350 (1965).
7. Gilmozzi, R. *et al. Nature* **328**, 318–320 (1987).
8. Danziger, I. J. *et al. Astr. Astrophys.* **117**, L13–L16 (1987).
9. Vidal-Madjar, A. *et al. Astr. Astrophys.* **177**, L17–L20 (1987).
10. Rybicki, G. B. & Lightman, A. P. *Radiative Processes in Astrophysics* (Wiley, New York 1979).
11. Landstreet, J. D. *Astrophys. J.* **258**, 639–650 (1982).
12. Koshida, M. *IAU Circ. No.* 4338 (1987).
13. Svoboda, R. *IAU Circ. No.* 4340 (1987).

Oxygen stoichiometry, superconductivity and normal-state properties of YBa₂Cu₃O_{7- δ}

R. J. Cava, B. Batlogg, C. H. Chen, E. A. Rietman, S. M. Zahurak & D. Werder

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Probably the most unusual crystal-chemical aspect of the new high-temperature oxide superconductors is their significant variability in oxygen content. For the YBa₂Cu₃O_{7- δ} family of compounds, the oxygen stoichiometry varies in the range $0 < \delta < 1$ under normal synthetic conditions, with the optimal 91-K superconductive transitions obtained near $\delta = 0$ (see, for example, ref. 1). Work presented so far on the dependence of both superconducting and normal-state properties on oxygen stoichiometry has considered only the value of δ in YBa₂Cu₃O_{7- δ} . Here we present data that indicate the issue to be considerably more interesting and complex, and propose that the microscopic variables thus far not considered include the distribution of oxygen over the available sets of sites in the structure, with that distribution strongly dependent on synthetic conditions. In addition to the dependence of the bulk transition temperature (T_c), normal-state resistivity and crys-

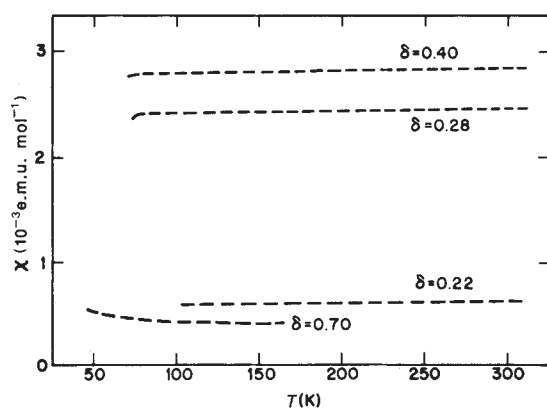


Fig. 1 Temperature dependence of magnetic susceptibility for $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ samples prepared by the gettered annealing technique.

tallographic symmetry on the microscopic distribution, we find the magnitude of the magnetic susceptibility in the normal state to be strongly affected by the microscopic state of the oxygen subsystem. We also report the existence of a single-phase 60-K bulk superconductor at stoichiometry $\text{YBa}_2\text{Cu}_3\text{O}_{6.6-6.7}$, which we postulate to have an ordered array of oxygen vacancies, and two distinct plateau regions in the variation of T_c with x which are either unobserved or only hinted at under other synthetic conditions.

The oxygen-deficient samples were prepared from single-phase, polycrystalline $\text{YBa}_2\text{Cu}_3\text{O}_7$ by a low-temperature gettered annealing technique². Pellets were sealed in evacuated quartz tubes with bright zirconium foil, annealed at temperatures between 360 and 520 °C for two days and air-quenched. For the sample sizes used, the amount of oxygen gettered from the $\text{YBa}_2\text{Cu}_3\text{O}_7$ is dependent on both the temperature of the anneal and the surface area of the zirconium foil. The technique results in extremely uniform oxygen-deficient samples and allows ample time for ordering of oxygen defects, if any is to occur. All samples were based on the same master batch of single-phase $\text{YBa}_2\text{Cu}_3\text{O}_7$ powder. Pellets were sintered at 1,000 °C in oxygen for 16 h, then annealed at 500 °C in oxygen for 16 h, and furnace-cooled to room temperature. Oxygen content after gettered annealing was determined by weight loss on heating in hydrogen to 1,000 °C. Other parts of the samples were used for powder X-ray diffraction, d.c. magnetization (SQUID magnetometer) and resistivity (four-probe a.c., 26 Hz) measurements.

Figure 1 shows the temperature dependence of the magnetic susceptibility (χ) in the normal state for selected oxygen-deficient samples prepared by gettered annealing. χ is either independent of temperature or decreasing with temperature for all δ up to 0.5, with only the slightest sign of a moment appearing at $\delta = 0.70$. This is in considerable contrast to our earlier published results on a $\delta = 0.1$ sample³, which showed a local moment of 0.9 Bohr magnetons (μ_B) per formula unit, and the Curie-Weiss behaviour of χ observed by many groups (see for instance refs 5-7). The local moment (ref. 4 and H. Katayama-Yoshida, personal communication) can either be intrinsic, or due to the presence of impurities in the samples. As our materials are phase-pure (by X-ray diffraction) at better than the 1% level, we believe that the presence or absence of the local moment is an intrinsic effect worthy of further study. The synthetic conditions described here for $\text{YBa}_2\text{Cu}_3\text{O}_7$ result reproducibly in material without a local moment.

For materials prepared by oxygen removal from $\text{YBa}_2\text{Cu}_3\text{O}_7$ by flowing gas⁵⁻⁷ at elevated temperatures, $\chi(T)$ is Curie-Weiss-like, with increasing local moment as a function of increasing oxygen deficiency or a δ -independent moment of $0.9 \mu_B \text{ f.u.}^{-1}$ for $0 < \delta < 0.5$ which then increases to $1.9 \mu_B \text{ f.u.}^{-1}$ at $\delta = 0.9$. As shown in Fig. 1, material can be prepared by the zirconium gettering technique at a stoichiometry of $\text{YBa}_2\text{Cu}_3\text{O}_{6.3}$ without significant local moments.

Table 1 Orthorhombic cell parameters and resistivities of $\text{YBa}_2\text{Cu}_3\text{O}_x$ at room temperature

x	a (Å)	b (Å)	c (Å)	ρ (mΩ cm)
7.0	3.822 (1)	3.891 (1)	11.677 (2)	1-1.5
6.78	3.827 (2)	3.895 (2)	11.722 (7)	5
6.76	3.830 (2)	3.898 (2)	11.728 (6)	4.9
6.72	3.835 (4)	3.890 (4)	11.716 (11)	3.8
6.60	3.831 (2)	3.889 (2)	11.736 (6)	3.5
6.53	3.838 (2)	3.887 (2)	11.747 (6)	4.7
6.51	3.845 (3)	3.887 (3)	11.768 (8)	5.4
6.30	3.851 (5)	3.883 (5)	11.789 (16)	9.2

Comparison of published data and those presented here reveals that specification of the total oxygen stoichiometry in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ is not sufficient to characterize the oxygen subsystem. We propose that the magnetic susceptibility is a sensitive measure of the local copper environment, especially with regard to the microscopic oxygen configuration. The variables thus far 'hidden' in the problem are the relative occupancies of the sites available for the oxygen within the cation array. Referred to the well-known $\text{YBa}_2\text{Cu}_3\text{O}_7$ structure (see, for example, ref. 8), the oxygen sites we consider relevant are the $(0, \frac{1}{2}, 0)$ site in the copper-oxygen chains, which would be fully occupied in an ideal structure, the $(\frac{1}{2}, 0, 0)$ site between chains, completely vacant in the ideal structure, and the $(\frac{1}{2}, 0, \frac{3}{8})$, $(0, \frac{1}{2}, \frac{3}{8})$ sites in the copper-oxygen planes, fully occupied in the ideal structure. Recent neutron diffraction work⁹ shows that $(0, \frac{1}{2}, 0)/(\frac{1}{2}, 0, 0)$ disorder occurs over a wide range of oxygen stoichiometry at temperatures above 450 °C. We propose that it is the relative occupancies of these sites that determine the normal-state magnetic susceptibility and resistivity, the crystallographic symmetry and unit-cell axis lengths, and the superconducting transition temperature; the configuration of the oxygen subsystem especially influences the magnetic state of the copper atoms.

The relative occupancy of the oxygen sites depends critically on the thermal history of the samples, and the final oxygen configuration of an oxygen-deficient sample may depend on the initial oxygen configuration. The $\text{YBa}_2\text{Cu}_3\text{O}_{6.3}$ sample prepared by gettered annealing from a $\text{YBa}_2\text{Cu}_3\text{O}_7$ sample with a well ordered oxygen array remains orthorhombic (see Table 1), probably because that initial ordering is at least partially maintained. $\text{YBa}_2\text{Cu}_3\text{O}_{6.3}$ prepared in the usual manner at high temperatures, where there is significant oxygen disorder, reflects that disorder, and is well into the compositional region of tetragonal symmetry.

Figure 2 shows the results of the resistive measurements of T_c for materials prepared by zirconium gettering as a function of oxygen stoichiometry. The 10-90% transition widths, indicated by the vertical bars, are quite narrow, except at the oxygen concentration where T_c is changing quickly. The narrow transition widths indicate that the oxygen in the samples is homogeneously distributed. The double plateau structure at 90 K and 60 K is dramatic and unmistakable. This plateau is either not observed, or only suggested, in materials prepared at higher temperatures by flowing gas or quenching techniques. The materials are all single-phase and orthorhombic. The amount of magnetic flux exclusion below T_c is equivalent for the 90-K and 60-K materials, indicating that the $\text{YBa}_2\text{Cu}_3\text{O}_{6.6-6.7}$ phase is also a good bulk superconductor. The resistivity of this compound shows a metallic temperature dependence. Further details of the characteristics of these materials will be reported elsewhere¹⁰; room-temperature resistivities are included in Table 1. Superconducting material with a bulk T_c of 55 K and a semiconducting resistivity characteristic has been prepared in the same stoichiometry range as our metallic 60-K material by a different synthetic technique¹¹; in our model of the importance of microscopic oxygen configuration, dependent on thermal history, such a difference is not unexpected.

Figure 2 also shows the value of the magnetic susceptibility at 100 K for the zirconium-annealed samples as a function of

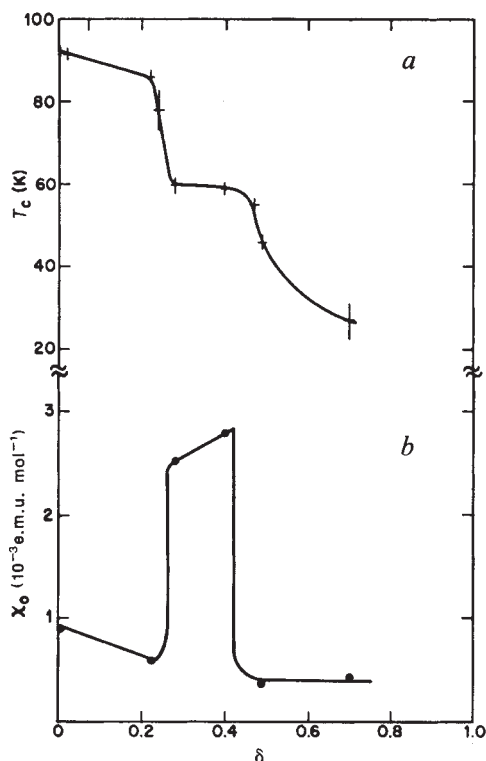


Fig. 2 *a*, Oxygen-content dependence of the resistive superconducting transition temperature in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. Vertical bars indicate the width of the transition from 90% to 10% of the normal-state resistivity, with the midpoint of the transition marked with a cross. *b*, Magnetic susceptibility at 100 K as a function of oxygen content in the same stoichiometry range.

oxygen stoichiometry. The data show a dramatic increase in susceptibility in the same composition region where the 60-K T_c plateau is observed. We do not have a detailed microscopic explanation for this non-monotonic behaviour of χ versus δ , but we suggest that in the range of δ in which χ is high, the oxygen subsystem, when annealed at the right temperature, assumes a microscopic ordering which leads to new short-range periodicities. Such order then produces new features in the energy-level scheme with a high density of states at the Fermi level.

Our results suggest that the value of the normal-state magnetic susceptibility, χ_0 , is very sensitive to the microscopic oxygen configuration, and that using it as a probe of the normal electron state requires consideration of the factors discussed here. Careful preparation and subsequent characterization are needed, to be able to draw meaningful conclusions.

We thank J. Remeika for describing the zirconium gettered annealing technique to us, and P. K. Gallagher for helpful discussions.

Received 20 August; accepted 15 September 1987.

- Gallagher, P. K., O'Bryan, H. M., Sunshine, S. A. & Murphy, D. W. *Mater. Res. Bull.* **22**, 995-1006 (1987).
- Chailout, C. & Remeika, J. P. *Solid St. Commun.* **56**, 833-836 (1985).
- Cava, R. J. *et al. Phys. Rev. Lett.* **58**, 1676-1679 (1987).
- Muller, J. *Pap. presented at Eur. Workshop High T_c Superconductors and Potential Applications*, Genoa, 1-3 July (1987).
- Johnston, D. C., Jacobson, A. J., Newsana, J. M., Lewandowski, J. T. & Goshorn, D. P. *Pap. presented at Am. chem. Soc. Meet. New Orleans* (31 August-4 September, 1987).
- Takagi, H. *Pap. presented at Workshop High T_c Superconductors and Potential Applications*, Genoa, 1-3 July (1987).
- Tarascon, J. M. *et al. Pap. presented at Am. chem. Soc. Meet. New Orleans* (31 August-4 September, 1987).
- David, W. I. F. *et al. Nature* **327**, 310-312 (1987).
- Jorgensen, J. D. *et al. Solid St. Commun.* (in the press).
- Cava, R. J. *et al. Phys. Rev. B* (in the press).
- Tarascon, J. M. *et al. Proc. Spring Meet. Mater. Res. Soc.* 21-24 April (eds Gubser, D. V. & Schluter, M.) (Mater. Res. Soc., in the press).

Detection of oxygen ordering in superconducting cuprates

A. Ourmazd* & J. C. H. Spence†

* AT&T Bell Laboratories, Holmdel, New Jersey 07733, USA

† Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

The recent discovery of superconductivity in ceramic cuprates¹ continues to lead to spectacular increases in the critical temperature. Results from several neutron diffraction experiments have shown that the system $\text{YBa}_2\text{Cu}_3\text{O}_7$ displays long-range order in the disposition of its oxygen vacancies, giving rise to two- and one-dimensional features in the structure^{2,3}, which may have important electronic implications^{4,5}. Although diffraction techniques have proved invaluable in determining the overall structure, the local structure should also be investigated on a microscopic level. Because of the relatively small scattering power of the light oxygen atoms, lattice images obtained from very thin films of the new superconductors are insensitive to the precise arrangement of the oxygen vacancies. It is known however, that due to dynamical effects, the thickness dependence of lattice images can be extremely sensitive to small effects, such as ionicity⁶. This can be exploited to detect the presence of long-range order in the oxygen vacancies⁷. Other investigators have produced experimental⁸ or simulated⁹ images, which also purport to provide evidence for the presence or detectability of order. Indications of superconductivity at very high temperatures in multiphase materials¹⁰, and the potential importance of determining the oxygen arrangement in small regions of a multiphase sample, mean that it is important to establish in detail the conditions needed for a reliable microscopic investigation of the presence of long-range order in these materials. Here we present theoretical and experimental evidence that, under a carefully chosen and controlled set of experimental parameters, it is possible to detect the presence of long-range order in the disposition of the oxygen vacancies by high-resolution transmission electron microscopy.

We study the lattice imaging of $\text{YBa}_2\text{Cu}_3\text{O}_7$ at 400 kV and discuss the oxygen arrangement in the basal plane. Similar arguments apply to the oxygen vacancy at the Y plane and to imaging at lower accelerating voltages. We assume that no systematic changes occur in the metal-frame stoichiometry with increasing sample thickness. Our simulations¹¹ show that the small difference between the lengths of the *a* and *b*-axes has no significant effect on the images. Assuming complete ordering of the oxygen vacancies as shown in Fig. 1, the problem of the detection of order is equivalent to a distinction between the [100] projection, when the O-Cu-O chains lie perpendicular to the electron beam, and the [010] projection, when the chains are in the direction of the beam. Figure 2 shows the Pendellosung plots for these two projections. (SHRIL suite of multi-slice programs¹¹, 128 square array, acc. voltage: 400 kV, C_s : 1 mm, Δ : 100 Å, beam divergence: 1 mr, objective aperture: 1.6 Å^{-1} . The simulations do not depend sensitively on the choice of parameters.) Significant differences in the beam phases and amplitudes only begin to appear at thicknesses $\geq 80 \text{ Å}$. This implies that a distinction between the [100] and [010] projections can only be made reliably in thicker regions of the crystal. Without information about the presence or degree of order, it is also necessary to distinguish between the ordered and disordered configurations. Figure 3 shows a series of image simulations carried out for a sample with various degrees of order in the [100] and [010] orientations. The defocus value shown is -500 Å , which produces high-contrast images easily recognizable in an experiment. Similar simulations have been made for $-300 \leq \Delta f \leq -900 \text{ Å}$. These clearly show that it is difficult, if not impossible, to distinguish reliably between the ordered [100]

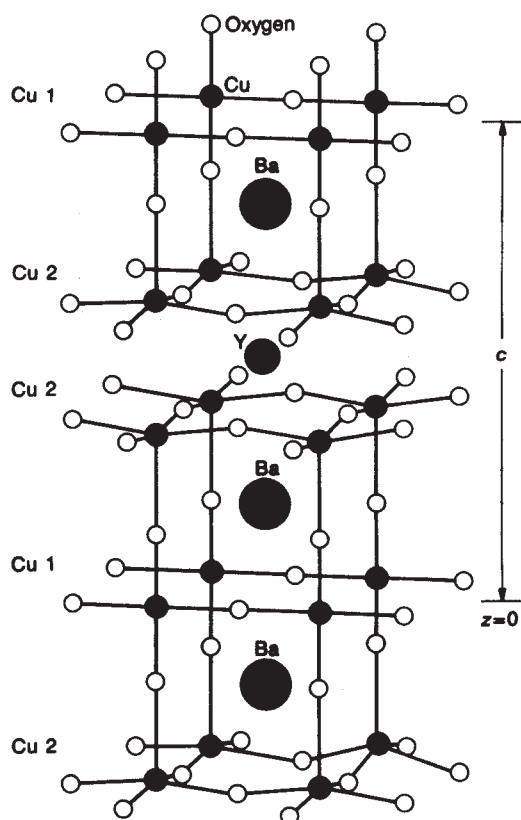


Fig. 1 Schematic representation of the structure of $\text{YBa}_2\text{Cu}_3\text{O}_7$.

and disordered configurations, essentially at any defocus or thickness. Even when the degree of order is less than one, however, the ordered [010] arrangement is distinct both from the [100], and the disordered arrangement, provided images are obtained from regions of sufficient thickness ($\geq 100 \text{ \AA}$).

In the [010] projection, when the O-Cu-O chains lie in the direction of the beam, the presence of order gives rise to a characteristic strong feature at the basal plane, which becomes a continuous white line with increasing thickness. Our simulations show that, for $-300 < \Delta f < -600 \text{ \AA}$ and $90 < t < 135 \text{ \AA}$, the presence of this continuous line constitutes a unique fingerprint for the presence of O-Cu-O chains parallel to the electron beam. The key to the detection of oxygen ordering thus lies in the observation of a series of images as a function of thickness, similar to those shown in Fig. 3c and d. These are readily distinguishable from the ordered [100] and disordered configurations. We find, moreover, that oxygen desorption from the material (the uniform removal of oxygen from all basal sites, reducing the stoichiometry to $\text{YBa}_2\text{Cu}_3\text{O}_6$) also destroys the continuous white line, the fingerprint of the ordered [010] arrangement (Fig. 3e). The quantification of the degree of order (a distinction between rows c and d) is likely to be experimentally difficult and may be unreliable (see below).

Although the simulations of Fig. 3 show clear differences between the different oxygen configurations under ideal conditions, they place severe restrictions on the experimental conditions to be achieved before a distinction can be made. We believe that misalignments such as astigmatism can be detected with sufficient accuracy by optical diffractometry. But small tilt misalignments are difficult to detect and can seriously modify lattice images obtained from thicker crystals, and the question arises whether such misalignments render the detection of oxygen ordering impractical or unreliable. Figure 4 shows a series of simulations for the [100] and [010] projections as a function of tilt misalignment. (We assume that the major effect is due to small errors in aligning the crystal orientation with respect to

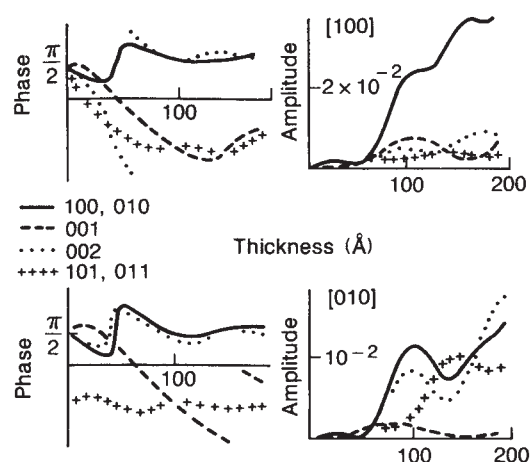


Fig. 2 Pendellosung plots showing the thickness dependence of various beams for the [100] and [010] projections in the presence of perfect oxygen order. Phases and amplitudes are referred to those of the undiffracted beam. Note that significant differences appear at thicknesses $\geq 80 \text{ \AA}$. At such thicknesses, the behaviour of the beams depends rather sensitively on the oxygen configuration and order parameter.

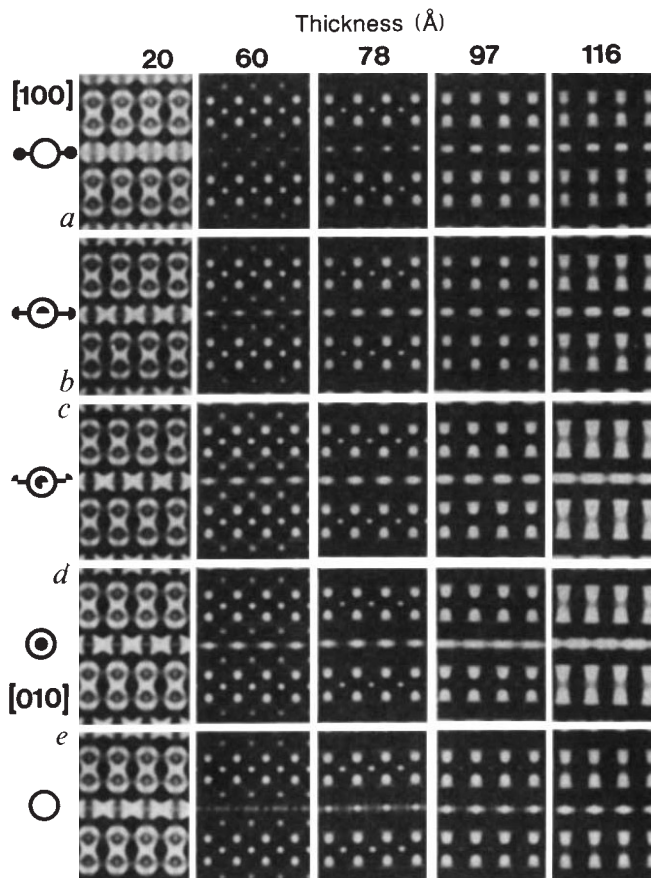


Fig. 3 Simulated lattice images at different foil thicknesses for various oxygen configurations. Unless stated, the usual orthorhombic lattice parameters are used. Defocus, $-500 \text{ \AA}$; a, [100] projection, perfect order; b, [100] (or [010]) projection in the absence of order ($a=b=3.87 \text{ \AA}$); c, [010] projection with 75% of the oxygens lying along b and 25% lying along a, producing a degree of the order of 0.5; d, [010] projection, perfect order; e, [100] or [010] projection, $\text{YBa}_2\text{Cu}_3\text{O}_6$ ($a=b=3.87 \text{ \AA}$). Note that rows a, b and e are similar, but that for foil thicknesses $\geq 100 \text{ \AA}$, the continuous white line between the basal planes distinguishes rows c and d, acting as a fingerprint for the presence of order in the [010] projection.

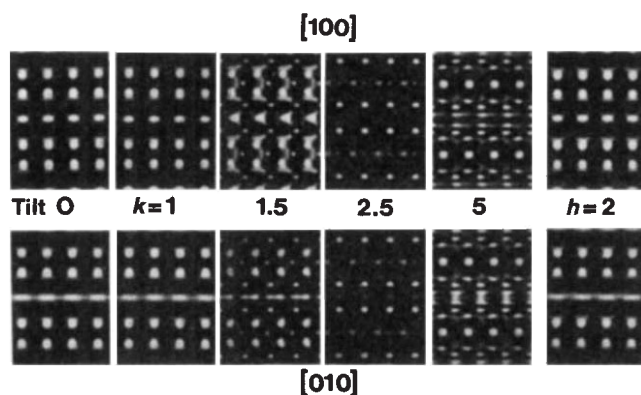


Fig. 4 Effect of tilt misalignment on lattice images in the [100] and [010] projections in the presence of perfect order. The centre of the Laue circle is at (h, k) . Only the index specified has been changed from zero for each simulation; $h = 1$ corresponds to a tilt of ~ 4.3 mrad, with the centre of the Laue circle displaced towards the (100) or (010) spot, $k = 1$ to ~ 1.4 mrad, with the centre displaced toward the (001) spot. Beam divergence, 1 mrad; defocus, -500 Å; sample thickness, 100 Å. Note that tilt misalignments can remove the differences between the two projections (and also the ordered [010] and disordered configurations), but cannot mimic the effect of order in the [010] projection.

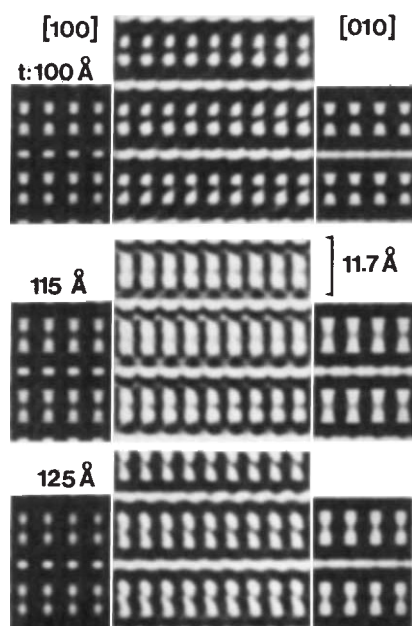


Fig. 5 Experimental images (centre) and simulations at three different thicknesses. Defocus, -500 Å; perfect oxygen order and alignment assumed in the simulations. It is likely that, experimentally, small tilt misalignments (≤ 1.2 mrad) may be present, or that the degree of order is < 1 . The experimental images nevertheless match the [010] ordered simulations much more closely and thus indicate the presence of order (of O-Cu-O chains parallel to the electron beam).

the electron beam direction. There are well-established procedures for the accurate alignment of the beam with respect to the optic axis.) The amount of tolerable misalignment depends on the sample thickness. For a film thickness of ~ 100 Å, misalignments greater than ~ 2.5 mrad, displacing the centre of the Laue circle towards the (001) spot, remove the differences between the [100] and [010] projections, and also between the ordered [010] and disordered configurations. For tilts displacing the centre of the Laue circle towards the (100) or (010) spots, however, larger deviations from exact alignment can be tolerated before the evidence for order is lost. Thus, tilt misalignments

can change the character of the images and remove the possibility of detecting the presence of order, but, significantly, do not mimic the effect of order. Thus the experimental observation of images of the type shown in Fig. 3c and d provides important evidence for the absence of significant misalignment or oxygen desorption, as well as the presence of oxygen ordering.

In Fig. 5 we present experimental data on the thickness dependence of lattice images in regions with thicknesses and orientation appropriate for the detection of order. These images were obtained in a single exposure of a negative at the same defocus of -500 Å. The experimental defocus was first deduced to within $\sim 10\%$ by optical diffractometry on the neighbouring amorphous regions of the sample and further refined by fitting of simulated images to within ~ 20 Å. The thickness assignments were deduced by image matching to within 8 Å and checked by cross-calibration with the Pendellosung oscillations. The experimental images fit the ordered [010] images much more closely, and thus provide evidence for the presence of long-range order in the oxygen arrangement in the basal planes. By obtaining convergent beam electron diffraction patterns of this area and the neighbouring twins, we have independently established the projection to be [010], thus confirming the validity of our approach. In general, the white lines observed separating the basal planes show a modulation somewhat larger than present in the simulations. This could be due to the presence of an order parameter of less than one, produced during growth, or subsequently, due to oxygen desorption, and/or disordering during sample preparation. Additionally, or instead, small tilt misalignments may be responsible for the observed modulations of the white line. These observations emphasize the difficulty of quantifying the degree of order by lattice imaging, which might depend on an accurate determination of this modulation.

We thank D. W. Johnson Jr for providing the samples used in this study. The work of one of us (J.C.H.S.) was supported in part by the NSF.

Received 22 June; accepted 12 August 1987.

1. Bednorz, J. G. & Müller, K. A. Z. *Phys.* **B64**, 189 (1986).
2. Capponi, J. J. *et al. Europhys. Lett.* **3**, 1201-1308 (1987).
3. David, W. I. F. *et al. Nature* **327**, 310-312 (1987).
4. Matthiess, L. F. & Hamann, D. R. *Solid State Commun.* (submitted).
5. Yu, J., Freeman, A. J. & Xu, J. H. *Phys. Rev. Lett.* **58**, 1035-1037 (1987).
6. Anstis, G. R., Lynch, D. F., Moodie, A. F. & O'Keefe, M. A. *Acta crystallogr.* **A29**, 138-147 (1973).
7. Ourmazd, A. *et al. Nature* **327**, 308-310 (1987).
8. Shaw, T. *Physics Today* **40**, 17 (1987).
9. Hewat, E. A. *et al. Nature* **327**, 400 (1987).
10. Ovshinsky, S. R. *et al. Phys. Rev. Lett.* **58**, 2579-2581 (1987).
11. O'Keefe, M. A., Busek, P. R. & Iijima, S. *Nature* **274**, 322-324 (1978).

Experimental evidence for convective rolls in finite two-dimensional molecular models

M. Mareschal & E. Kestemont

Faculté des Sciences, Code Postal 231, Université Libre de Bruxelles, Boulevard du Triomphe, B1050, Bruxelles, Belgium

Molecular dynamics (MD) techniques, based on explicit numerical solution of the equations of motion of the particles in a molecular system, have been successful in the study of models simulating fluids in equilibrium configurations as well as in stationary states due to the presence of external constraints. The possibility of observing true hydrodynamical behaviour has however been frequently questioned. There has been much recent interest¹ in new computational techniques which might be used to approach the difficult problems of flow and convection². Results from studies of cellular automata^{3,4} are very promising. Fluid flows obstructed by plates or cylinders have been studied, using MD techniques, by Rapaport and Clementi⁵, who demonstrated the appearance of

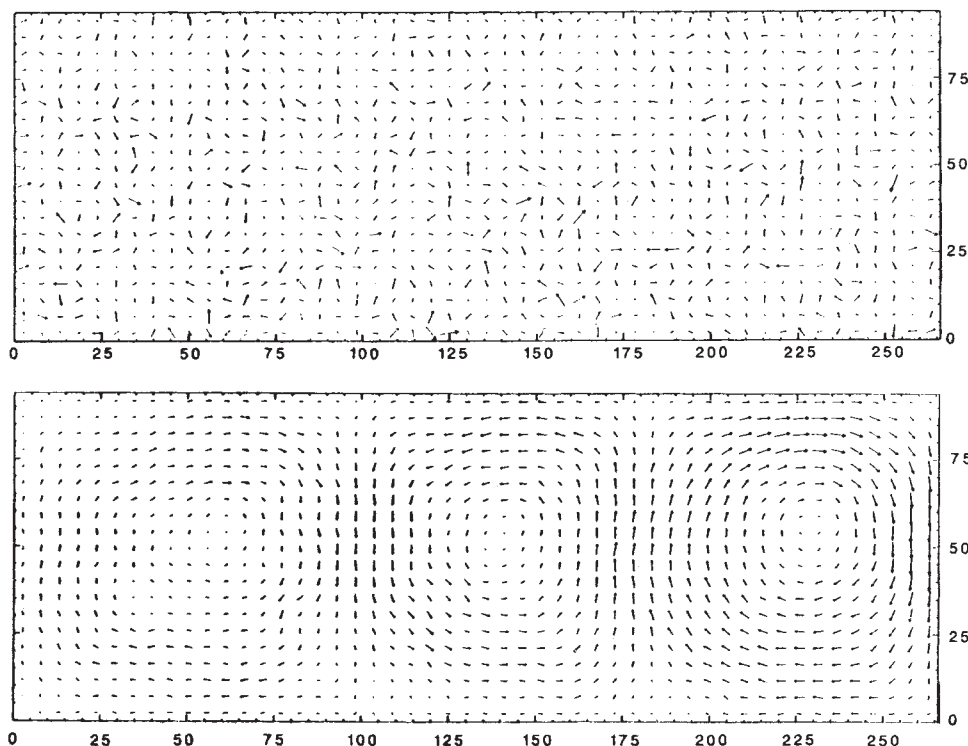


Fig. 1 Initial streamlines in a finite two-dimensional system of 5,040 hard disks under the influence of a thermal gradient and an external force. Number density is 0.2 and g is 0.09. Density velocities are normalized.

Fig. 2 Same system as in Fig. 1 after 13 million collisions. Rayleigh number is estimated to be 1,780. Each arrow is a time average of the local velocity density over the entire run.

vortices, although it was generally argued that such simulations were out of range of the power of present computers. Here, we study finite systems of hard disks⁶ in the presence of both a thermal gradient and an external constant force. For well-chosen geometrical parameters, (density, gradient value and magnitude of the applied force), we observe the appearance of convective flow in these very small two-dimensional systems. Starting from a completely chaotic state, the system evolves to an ordered state after a few million collisions between particles. The vortices remain stable thereafter for at least a few thousand collisions per particle.

Our system consists of 5,040 hard disks placed in a rectangular box (L_x, L_z) at a number density of 0.2, the unit of length being the diameter of a disk. The dimensions of the box are respectively 266 in the x direction and 94 in the z direction. The $z=0$ boundary represents a reservoir at temperature $T_0 + \Delta T$ while at $z=L_z$ temperature is maintained at T_0 . Initially, the particles are placed at random in the system, their velocity components being chosen from equilibrium distributions at the local temperature extrapolated from a linear profile between the temperatures of the reservoirs. The boundaries of the system act as walls. On the x boundaries, particles undergo simple specular reflections. When a particle hits a thermal wall, its new velocity is sampled from the local distribution but the sign of the tangential (x) component is preserved. A constant external force acting in the opposite direction of the thermal gradient is applied on the molecules. In our simulations, the temperatures have been chosen as 1.0 and 10.0 respectively, and the constant external force corresponds to a gravitational acceleration of 0.09 in units where the mass of the particles, the molecular diameter and the Boltzmann constant k_B are 1.0. For the given number density, the measured temperature at $z=L_z/2$, and using the Enskog values⁷ for the kinematic viscosity and thermal diffusivity, we estimate the Rayleigh number to be 1,780.

The system evolves in time according to our standard computer program, incorporating the principles described in ref. 8. To calculate local properties, the system is divided in 20×50 boxes, and from the chosen random initial condition is followed in phase space. After about one million collisions between particles (nearly 400 collisions per particle) we begin to compute at equal time intervals various mean quantities, for example,

the local temperature (defined as the mean kinetic energy of the particles in the given box), the mean number density and the mean velocities of the particles.

We give in Fig. 1 a sketch of the initial mean velocity densities over a time interval equal to the mean collision time. As expected, mean velocities are random, and of the order of the square root of the local temperature. Figure 2 represents the same mean velocities, but after an evolution time of the order of a few thousand collision times. Vortices corresponding to a regular flow of particles from the warm bottom plate to the cold top one can now be seen. In this case, the mean velocities are one order of magnitude lower than the thermal velocities. Note that the linear temperature profile between the two plates is still present. Figure 3 gives this profile together with the variation of the local number density and local pressure in the system as functions of position. As time goes by, the structure becomes more and more pronounced, the asymmetry between the left and right sides of the system disappearing slowly. The vortices

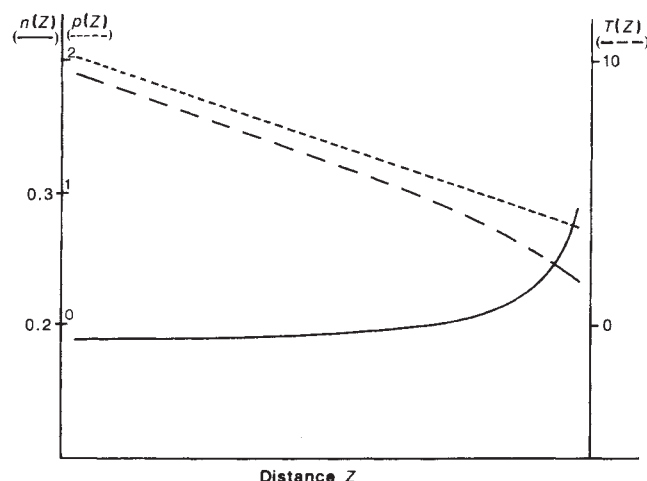
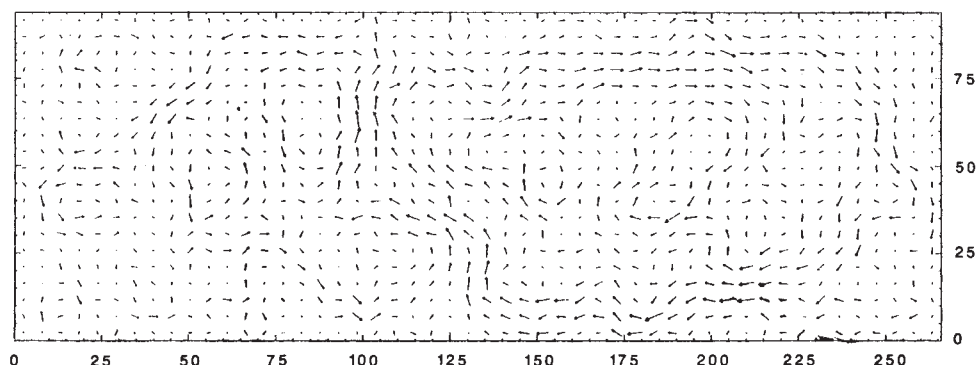


Fig. 3 Temperature, density and pressure profiles in the same system as a function of distance from the hot plate ($z=0$).

Fig. 4 Streamlines in a system of 5,040 particles with the same geometry and density as in Fig. 1, but with different values of the thermal gradient and of the gravitational acceleration, giving a Rayleigh number equal to 990.



remained stable throughout the experiment, which lasted for more than twenty million collisions (the computer program needed nearly 400 h of central processor time on a 4341 IBM computer); this corresponds to four times the time needed for heat to diffuse over a distance equal to L_z .

What is most surprising is the existence and persistence of such structures in very small systems. Most physicists were convinced of the need for systems two or more orders of magnitude bigger than this to observe such behaviour. It thus appears that the hypothesis of local equilibrium is verified on the microscopic level⁸ and also that true collective motions exist for long, hydrodynamic times at dimensions close to a mean free path. The choice of the molecular and external parameters is crucial, as can be seen in Fig. 4, where we have drawn the flows for a system with the same number of particle and the same geometry, but with slightly different external force and temperature gradient. In this case, the Rayleigh number is reduced to 990, and, while whirlpools seem indeed to be present, there are no stable ordered structures. In particular, the computed histories of these model fluids show that the onset of an ordered state results from competition between spontaneous fluctuations giving rise to small structures and damping caused by their

environment. Such a description falls outside the range of hydrodynamic equations.

We acknowledge constant encouragement from Professor I. Prigogine and stimulating discussions with Professor G. Nicolis and Drs M. Malek Mansour and A. Garcia. We thank Professor E. Clementi for his hospitality at IBM Corporation (Kingston, New York), where most of the preliminary computations have been performed on his experimental LCAP-IBM computer. The results presented here we obtained with the help of the Belgian Fonds National de la Recherche Scientifique at the FNRS-IBM computer centre of Namur. IBM-Belgium and Instituts Internationaux de Physique et de Chimie fondés par E. Solvay are thanked for financial assistance.

Received 27 February 1987; accepted 16 July 1987.

1. Hillis, W. D. *Nature* **325**, 299–300 (1987).
2. Evans, D. J. & Hoover, W. G. *A. Rev. Fluid Mech.* **18**, 243–264 (1986).
3. Wolfram, S. *Nature* **311**, 419–424 (1984).
4. Frisch, U., Hasslacher, B. & Pomeau, Y. *Phys. Rev. Lett.* **56**, 1505–1508 (1986).
5. Rapaport, D. D. & Clementi, E. *Phys. Rev. Lett.* **57**, 695–698 (1986).
6. Erpenbeck, J. J. & Wood, W. W. in *Statistical Mechanics B* (ed. Berne, B. B.) Ch. 1 (Plenum, New York, 1977).
7. Gass, D. M. *J. chem. Phys.* **54**, 1898–1902 (1971).
8. Mareschal, M. & Kestemont, E. *Phys. Rev. A* **30**, 1158–1160 (1984).

Mantle-derived helium in two Peruvian hydrothermal ore deposits

Stuart F. Simmons*[‡], F. J. Sawkins* & D. J. Schlutter†

* Department of Geology and Geophysics, † Department of Physics, University of Minnesota, Minneapolis, Minnesota 55455, USA

Strong differences exist between the helium isotopic compositions of mantle-derived magmas^{1–8}, axial vent hot-spring fluids^{9–14}, subduction-related magmas^{15–20} and many associated hot springs^{21–24}, and the atmospheric and crustal values (Fig. 1). Here we report helium isotope ratios for two late Tertiary fossil hydrothermal systems responsible for precious-metal, base-metal and tungsten mineralization, and compare them to those of circum-Pacific hot-spring emanations. The $^3\text{He}/^4\text{He}$ ratios of inclusion fluids hosted by gangue and sulphide minerals from Casapalca and Pasto Bueno, Peru, all lie between two and three times the atmospheric value, indicating the presence of a dilute component of mantle helium consistent with previously reported stable isotope evidence for the presence of magmatic fluids during mineralization^{25–27}. Although post-trapping modification of the helium isotope composition is possible, we suggest that these apparently low values represent the composition of helium inherited from contributing magmas. As a tracer of magmatic fluids in hydrothermal systems, helium isotopes are decoupled from the factors that govern other stable isotope signatures.

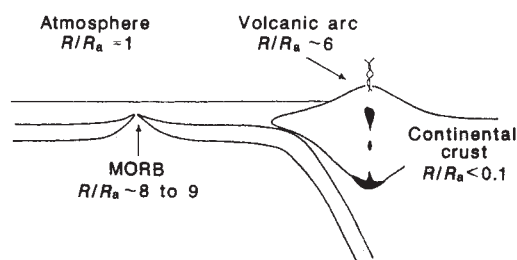


Fig. 1 Summary of helium isotopic compositions from various tectonic environments³⁶. MORB, mid-ocean-ridge basalts.

Of particular interest are the many subaerial, circum-Pacific hot springs that exhibit strong mantle helium signatures, with the $^3\text{He}/^4\text{He}$ ratio (R) greater than six times the atmospheric ratio (R_a), but which have O and H isotope signatures representative of local meteoric water; for example, the Broadlands geothermal field, New Zealand^{23,28}. For these hydrothermal systems, identification of mantle helium is evidence for the presence of a magmatic heat source at depth²³. Crystalline igneous rocks are unlikely to contain sufficient amounts of helium available for hydrothermal leaching as most helium would have partitioned favourably into melt and gas phases during crystallization, for example, mantle helium signatures in young oceanic basalts are obtained from glassy rims of pillows, whereas the crystalline cores contain seawater (atmospheric) helium^{2,29}.

We have investigated the $^3\text{He}/^4\text{He}$ ratios in fluid inclusions from two well studied Andean metal deposits, Casapalca and

[‡] Present address: Department of Geology, University of Auckland, Private Bag, Auckland, New Zealand.

Table 1 Temperature, salinity and stable isotope summary of mineralizing fluids

	Temperature (°C)	NaCl (% equiv. wt)	δD	$\delta^{18}O$	$\delta^{34}S$	$\delta^{13}C$
Casapalca	280–370	4–40	–48––60	4.0–9.5	0.5–3.5	–6.5––5.4
Pasto Bueno	175–290	2–17	–29––88	0.0–7.8	–5.48–2.85*	–4.1––11.9

From refs 25–27. * Pyrite data only.

Pasto Bueno, and these measurements add to their extensive fluid inclusion and isotopic databases^{25–27}. Both deposits are closely associated with subduction-related magmatism, but are quite different in terms of their contained metals, host rocks and genesis. At Casapalca, steeply dipping Ag–Pb–Zn–Cu-mineralized veins crosscut a folded sequence of Tertiary sedimentary and volcanic units. Only minor amounts of intrusive rocks can be mapped at the surface in the district, but the occurrence of contact-metamorphosed and metasomatized rocks at depth below the vein system indicates the presence of an underlying igneous body. This interpretation is further confirmed by the clear magmatic signatures of the mineralization process provided by the light stable isotope data (Table 1). The final stages of mineralization were characterized by an influx of meteoric waters which deposited calcite.

The tungsten-base-metal-bearing veins at Pasto Bueno crosscut the contact between the Consuzo Stock, a late Tertiary quartz monzonite, and adjacent Mesozoic sedimentary rocks. An approximate age of 8 Myr is indicated from K–Ar dating of hydrothermal sericite. O, H, C and Sr isotope studies indicate: (1) fluorite was deposited from fluids of magmatic origin; (2) tungsten of magmatic derivation was deposited during influx of meteoric fluids; and (3) sulphide mineralization occurred in the presence of magmatic, meteoric and a third evolved water of unknown provenance^{26,27}.

Helium isotopic compositions of inclusion fluids trapped in quartz, sphalerite, tetrahedrite and fluorite range between two and three times the atmospheric ratio for both deposits (Table 2; Fig. 2). These measurements indicate that a portion of the helium in inclusion gases is mantle-derived and are consistent with previously cited stable isotope evidence for the presence of magmatic fluids during mineralization. However, the figures appear low, considering both the strong magmatic stable isotope signature for both deposits, and the high R/R_a values for many circum-Pacific hot springs that are clearly dominated by meteoric

water. The possibility that helium isotope values are the result of lithium–nuclear reactions is discounted by the absence of any known lithium-bearing mineral in the vicinity of samples studied. Cosmogenic helium produced by spallation reactions at high elevations recently identified as a surficial phenomenon in the summit lavas at Maui^{30,31} and in the Fresno District, Mexico³² is also considered an unlikely source of 3He as all samples were taken from underground workings, at depths greater than 100 m below the surface.

Possible explanations for the relatively low R/R_a values we report include: (1) preferential diffusive loss of 3He from the fluid inclusions; (2) *in situ* generation of 4He due to radioactive decay of U and Th which may be present as dissolved constituents in the host minerals or impurities in inclusion fluids; (3) fractionation of 3He into the vapour phase during boiling, resulting in lowered concentrations of 3He in liquid-rich fluid inclusions; or (4) dilution of mantle helium with crustal helium. This latter would be due to possible mixing with meteoric fluids in the hydrothermal system, magmatic assimilation of crustal material, or magma 'ageing' where radiogenic helium accumulates due to a long residence time within the magma chamber³³.

Diffusive loss is very difficult to evaluate and the problem may be compounded further if trace amounts (less than 1 p.p.m.) of U and Th are present in host minerals. Craig and Lupton³ discussed these problems at length with respect to lower helium isotopic compositions in older basalt glass rims from pillow basalts. Although helium diffusivities in vesicular basalt glass and a mineral containing fluid inclusions are probably different their analysis shows that the residual R/R_a decreases logarithmically with respect to helium concentration. Only after more than 95% of the helium has been lost is the R/R_a half its inherited value, and thereafter the residual R/R_a becomes very sensitive to change in helium concentration. Accumulation of *in situ* radiogenic helium further enhances this sensitivity, therefore significant modification of helium concentration from the

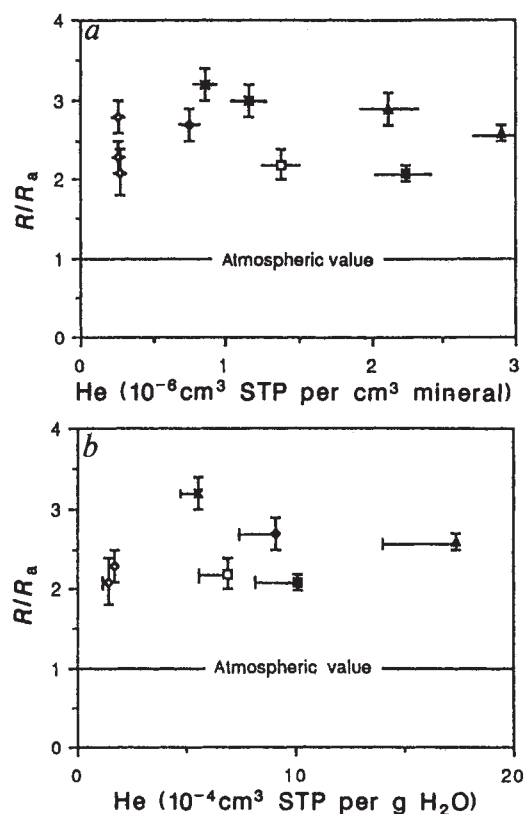
Table 2 Summary of temperature, $\delta^{18}O$, δD , $\delta^{34}S$, $^{87}Sr/^{86}Sr$ and $^3He/^4He$ (R/R_a) data of inclusion fluids from Casapalca²⁵ and Pasto Bueno^{26,27}, Peru

Location and sample	Temperature (°C)	$\delta^{18}O$	δD	$\delta^{34}S$	$^{87}Sr/^{86}Sr$	$R/R_a \pm 1\sigma$	$He(\times 10^{-6}$ cm ³ STP)	H ₂ O (mg)
Casapalca								
68-S-18 qtz (17 g)	359 (P), 336–356 (Ps)	3.8–5.4*				2.2 ± 0.1	9.0	13
68-S-30 sph (36 g)	354 (P), 323–333 (Ps)	7.3	–49––65			2.1 ± 0.3	2.4	16.6
68-S-66 qtz (12 g)	307–327 (P & Ps)					2.3 ± 0.1	—	
69-S-7 sph (21 g)	331 (P), 300–303 (Ps)		–53	2.4		2.8 ± 0.2	1.3	
69-S-22 sph (24 g)				1.76		2.3 ± 0.2	1.5	8.7
Pasto Bueno								
9H-148A wh fluor (19 g)	168 (P)	1.37*	–45.6			3.2 ± 0.2	5.1	9.2
9H-124 qtz (8 g)	150–180 (P)	2.20*	–35.8			2.1 ± 0.1	6.9	6.8
9H-150 fluor (14 g)	167–183 (P)		–46.8		0.70705	3.0 ± 0.2	5.1	
9Mn-126 tetr (22 g)				2.83		2.6 ± 0.1	12.5	7.2
9Mn-127 sph (31 g)				2.85	0.71228	2.7 ± 0.2	5.7	6.3
9H-152 tetr (20 g)			–113.1	–1.37		2.9 ± 0.2	8.3	
Blank						6.8 ± 5.8	0.075	

P, primary inclusions; Ps, pseudosecondary inclusions; qtz, quartz; sph, sphalerite; tetr, tetrahedrite; fluor, fluorite; wh, white.

* Calculated $\delta^{18}O$ composition of inclusion water.

Fig. 2 Plots of R/R_a versus helium concentration treated either with respect to host mineral volume (a) or with respect to water released from fluid inclusions (b). Helium isotope compositions were measured from fluids trapped in 99% pure, mineral separates of quartz (qt), sphalerite (sp), tetrahedrite (td) and fluorite (fl). Samples were initially cut to less than 1-cm diameter with a water-cooled saw, ultrasonically cleaned in acetone and placed in the sample chamber where they were evacuated to less than 10^{-4} atm overnight. The following day, they were heated to $>500^\circ\text{C}$ for 30 min to release gases by decrepitation of fluid inclusions. Helium and neon were separated from the other gases by a liquid N_2 -cooled, activated charcoal trap and collected in a glass bulb. The bulb was immediately transferred to the inlet line of a 3.75-in radius double-focusing Mattauch-Herzog geometry mass spectrometer³⁷, dedicated to terrestrial samples only. The sample gas was further purified by a second liquid N_2 -cooled, activated charcoal trap and a SAES getter, and then introduced into the mass spectrometer where ^3He and ^4He were measured simultaneously. Atmospheric standards were run before and after sample runs and averaged together to determine R/R_a reported in Table 2. Absolute amounts of helium reported were determined by comparing peak heights of known concentrations of helium in the air standards with the sample peak heights and are considered accurate to $\pm 10\%$. Quantities of water released during decrepitation of samples were measured by weight difference after the water had been collected in a liquid N_2 -cooled glass capillary. These are minimum values and are no more than 25% low. A procedural blank yielded $7.5 \times 10^{-8} \text{ cm}^3 \text{ STP He}$ and $R/R_a = 6.8 (\pm 5.3)$. *In vacuo* crushing experiments, although not used in the measurement of these samples, have been performed on hydrothermal quartz from underground mine workings in the Fresno district, Mexico, and have shown that the helium isotopic composition of gases released by this technique are identical to helium ratios measured of gas released by the thermal decrepitation technique³².



original trapped concentration should be accompanied by noticeable changes in the R/R_a . Utilization of four different mineral species of different density, chemistry and crystallography suggest that diffusion rates, as well U and Th contents, for each of these minerals are different. All the Peruvian R/R_a values, however, fall within a narrow range despite the broad spread in helium concentration (Fig. 2), suggesting that diffusion and *in situ* ^4He production were not important in lowering the inherited R/R_a .

In situ radiogenic accumulation of ^4He could instead result from dissolved U within inclusion fluids. Although we lack data on U concentration in the inclusion fluids, a calculation can be made which gives a maximum ^4He production in 8 Myr from an assumed initial U concentration and the following equations³:

$$J(^4\text{He}_R) = 0.2355 \times 10^{-6} \text{ U}^* \text{ cm}^3 \text{ g}^{-1} \text{ per } 10^6 \text{ yr}$$

where

$$\text{U}^* = \text{U}(\text{p.p.m.}) [1 + 0.123(\text{Th}/\text{U} - 4)]$$

Using a Th/U ratio of 0, given that thorium is highly insoluble³⁴, and a U concentration of 50 p.p.m. gives a total ^4He production of $4.8 \times 10^{-5} \text{ cm}^3 \text{ STP per g H}_2\text{O}$ in 8 Myr. Subtracting this amount of radiogenic helium from an intermediate helium concentration of $5.5 \times 10^{-4} \text{ cm}^3 \text{ STP per g H}_2\text{O}$ (sample 9H-148A) measured, gives an inherited helium isotopic composition of $R/R_a = 3.5$. This indicates that U concentrations far greater than 50 p.p.m. would be required to alter significantly the initial helium isotopic composition, a value that seems too high for fluids from ore-generating systems devoid of uranium minerals³⁵.

At Pasto Bueno, evidence of boiling exists only in the early stage of mineralization²⁶ and the minerals for this study came from later stage mineralization. Liquid-rich and vapour-rich inclusions were observed to coexist in Casapalca sphalerite²⁵ and these samples should provide a $^3\text{He}/^4\text{He}$ composition closer to that of fluids prior to boiling. Consequently, it seems unlikely that preferential vapour-phase partitioning of ^3He could account for the same low helium isotopic compositions in both deposits.

Dilution of mantle helium in the hydrothermal system by a meteoric fluid containing radiogenic crustal helium could, in part, account for lower R/R_a at Pasto Bueno; however, this is not consistent with the strong magmatic O and H signatures at Casapalca²⁵.

Finally, the helium isotope data from these Andean ore deposits could represent the actual helium isotopic composition of fluids emanating from cooling intrusions at depth, an explanation we favour. This interpretation would imply that mantle helium signatures were modified by incorporating radiogenic helium within the magma before helium was released into the hydrothermal system, a process that could involve either crustal assimilation or magma 'ageing'. The similar values for both Casapalca and Pasto Bueno would suggest similar degrees of contamination as the magmas rose through the thick Andean crustal section. If this mechanism is important in controlling helium isotope signatures in magmatic-hydrothermal systems then intraoceanic arcs with thinner crust may have higher R/R_a values. We have preliminary data from the veins in the Baguio Gold Camp, Luzon (Philippines) which average $R/R_a \approx 6$, supporting this hypothesis. Unfortunately, active geothermal systems in the Andes have not been investigated with respect to helium isotope geochemistry, thus precluding a regional study of the effect of thick continental crust on magmatic-arc helium ratios.

In conclusion, mantle-derived helium, albeit diluted, has been identified in fluid inclusions associated with arc-related fossil hydrothermal systems, implying communication with a cooling intrusive body during mineralization. Further work is planned to investigate more thoroughly the possible mechanisms leading to apparent lowering of mantle helium signatures. Nevertheless, the uniformity of our data from a variety of minerals in two rather different deposits, one of which exhibits stable and strontium isotope evidence for three distinct fluid provenances, suggests constancy of source for helium and a decoupling from hydrological factors. This interpretation is further reinforced by the circum-Pacific hot-spring helium data where stable isotope

evidence for meteoric waters is at variance with mantle helium isotope signatures.

We thank A. O. Nier in whose laboratory all helium isotope measurements were made, Gary Huss, E. Ito, A. O. Nier, S. Blake and P. R. L. Browne for critical evaluations of this manuscript, and E. C. Alexander for advice. This research was supported by NSF grant to F. J. Sawkins.

Received 28 May; accepted 3 August 1987.

- Krylov, A. Ya., Mamryin, B. A., Khabarin, L. A., Mazina, T. I. & Silin, Yu. I. *Geokhimiya* **8**, 1220-1225 (1974).
- Lupton, J. E. & Craig, H. *Earth planet. Sci. Lett.* **26**, 133-139 (1975).
- Craig, H. & Lupton, J. E. *Earth planet. Sci. Lett.* **31**, 369-385 (1976).
- Poreda, R. & Craig, H. *Eos* **60**, 969 (1979).
- Poreda, R., Craig, H. & Schilling, J. G. *Eos* **61**, 1158 (1980).
- Kurz, M. D. & Jenkins, W. J. *Earth planet. Sci. Lett.* **53**, 41-54 (1981).
- Kurz, M. D., Jenkins, W. J., Schilling, J. G. & Hart, S. R. *Earth planet. Sci. Lett.* **58**, 1-14 (1982).
- Kurz, M. D., Meyer, P. S. & Sigurdsson, H. *Earth planet. Sci. Lett.* **74**, 291-305 (1985).
- Lupton, J. E., Weiss, R. F. & Craig, H. *Nature* **266**, 244-246 (1976).
- Weiss, R. F., Lonsdale, P., Lupton, J. E., Bainbridge, A. E. & Craig, H. *Nature* **267**, 600-603 (1977).
- Lupton, J. E., Weiss, R. F. & Craig, H. *Nature* **267**, 603-604 (1977).
- Jenkins, W. J., Edmond, J. M. & Corliss, J. B. *Nature* **272**, 156-158 (1978).

- Lupton, J. E. *et al. Earth planet. Sci. Lett.* **50**, 115-127 (1980).
- Lupton, J. E. & Craig, H. *Science* **214**, 13-18 (1981).
- Baskov, Y. *et al. Geochim. Int.* **10**, 130-138 (1974).
- Craig, H., Lupton, J. E. & Horibe, Y. *Terr. Rare Gases, Adv. Earth planet. Sci.* **3**, 3-16 (1978).
- Craig, H., Lupton, J. E., Welham, J. A. & Poreda, R. *Geophys. Res. Lett.* **5**, 897-900 (1978).
- Nagao, K., Takaoka, N. & Matsubayashi, O. *Earth planet. Sci. Lett.* **53**, 175-188 (1981).
- Poreda, R. J. & Basu, A. R. *Earth planet. Sci. Lett.* **69**, 58-68 (1984).
- Sano, Y., Nakamura, Y., Wakita, H., Urabe, A. & Tominga, T. *Science* **224**, 150-151 (1984).
- Welham, J. A., Poreda, R., Lupton, J. E. & Craig, H. *Geothermics* **8**, 241-244 (1979).
- Torgersen, T. & Jenkins, W. J. *Geochim. cosmochim. Acta* **46**, 739-748 (1982).
- Torgersen, T., Lupton, J. E., Sheppard, D. S. & Giggenbach, W. F. *J. volcan. geotherm. Res.* **12**, 283-298 (1982).
- Sano, Y., Nakamura, Y. & Wakita, H. *Chem. Geol.* **52**, 1-8 (1985).
- Rye, R. O. & Sawkins, F. J. *Econ. Geol.* **69**, 181-205 (1974).
- Landis, G. P. & Rye, R. O. *Econ. Geol.* **69**, 1025-1059 (1974).
- Norman, D. I. & Landis, G. P. *Econ. Geol.* **78**, 451-465 (1983).
- Giggenbach, W. F. *N.Z. J. Sci.* **14**, 959-970 (1971).
- Dymond, J. & Hogan, L. *Earth planet. Sci. Lett.* **20**, 131-139 (1973).
- Kurz, M. *Nature* **320**, 435-439 (1986).
- Craig, H. & Poreda, R. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1970-1974 (1986).
- Simmons, S. F. thesis, Univ Minnesota (1986).
- Condomines, M. *et al. Earth planet. Sci. Lett.* **66**, 125-136 (1983).
- Rogers, J. J. W. & Adams, J. A. S. *Handbook of Geochemistry* (ed. Wedepohl, K. H.) (Springer, Berlin).
- Rich, R. A., Holland, D. A. & Petersen, U. *Developments in Economic Geology, Hydrothermal Uranium Deposits Vol. 6* (Elsevier, Amsterdam, 1977).
- Lupton, J. E. *et al. Rev. Earth planet. Sci.* **11**, 371-414 (1983).
- Nier, A. O. C. & Schlutter, D. J. *Rev. Sci. Instrum.* **56**, 214-219 (1985).

The relative importance of acidity sources for humic lakes in Norway

David F. Brakke*, Arne Henriksen & Stephen A. Norton†

Norwegian Institute for Water Research (NIVA), PO Box 333, Blindern 0314, Oslo 3, Norway

The possible role of natural processes contributing to lake acidification and the occurrence of acidic lakes ($\text{pH} < 5$) has received considerable attention¹⁻⁷, but there has been no direct test of the importance of some of the postulated natural processes in Scandinavia where large numbers of acidic lakes occur. Some authors have argued that lakes were acidic simply because of processes occurring in acid soils, and that the effects of acidic precipitation on these lakes were thereby overestimated^{1,2}. We present here an analysis of the role of strong and weak acids in the acidity status of moderately humic lakes (total organic carbon

(TOC) 2-15 mg l^{-1}) in two areas of Norway. The two areas received precipitation with differing acidity and amounts of sulphate. Our data suggest contributions by organic acids to acidity, but strong acids associated with sulphate further reduce the pH and explain the recent acidification of coloured lakes in southern Norway.

Dystrophic lakes, and especially bogs, may be naturally acidic due to the presence of organic acids⁸⁻¹¹. Organic anions contribute to the acidity of lakes¹², but the relative roles of strong mineral acids and acidity associated with organic anions in lakes in areas containing large numbers of acidic lakes, such as southern Norway, are poorly described. A notable exception is the demonstration that both anthropogenic sulphur and organic acids contribute to the acidity of surface waters in Nova Scotia¹². Palaeolimnological analyses based on diatom reconstructions have been presented as evidence for the acidification of lakes in Scotland, Norway and the northeastern USA associated with strong acids¹³⁻¹⁷. Diatoms have also been used to reconstruct historic concentrations of TOC; these results suggest some reduction in TOC during acidification¹⁵.

Here we have assessed the relative contributions of acidity associated with organic anions and sulphate. To do this, we selected and sampled 94 lakes from two areas of Norway draining peaty soils and known to contain coloured lakes. The lakes in each area were located on granitic bedrock, were of similar size and had similar ranges of concentrations of base cations

* Present addresses: Institute for Watershed Studies, Western Washington University, Bellingham, Washington 98226, USA and † Department of Geological Sciences, University of Maine, Orono, Maine 04469, USA.

Table 1 Minimum, mean and maximum values for selected chemical variables in lakes at Gulsvik (G) and Nord-Trøndelag (NT), Norway

Parameter	Gulsvik ($n = 48$)			Nord-Trøndelag ($n = 46$)		
	min.	mean	max.	min.	mean	max.
Conductance ($\mu\text{S cm}^{-1}$)	1.18	2.01	2.58	1.60	2.82	4.51
Ca (mg l^{-1})	0.33	1.20	2.65	0.27	1.74	4.24
Mg (mg l^{-1})	0.10	0.23	0.39	0.28	0.54	1.09
Na (mg l^{-1})	0.28	0.56	1.23	1.10	2.70	5.30
Cl (mg l^{-1})	0.30	0.59	1.00	1.50	3.96	6.00
SO ₄ (mg l^{-1})	2.50	3.31	4.50	1.20	1.69	3.10
NO ₃ -N ($\mu\text{g l}^{-1}$)	1.0	17.7	62	1.0	11.7	81
TOC (mg l^{-1})	1.86	9.1	14.3	2.2	8.1	15.2
Anion deficit ($\mu\text{equiv. l}^{-1}$)	-8.0	31.0	78	11.5	35.5	72.3
Ca* + Mg ($\mu\text{equiv. l}^{-1}$)	21.4	75.4	146	16.2	105	244
SO ₄ * ($\mu\text{equiv. l}^{-1}$)	50.3	67.2	91.6	11.0	23.6	54.6
Na* ($\mu\text{equiv. l}^{-1}$)	0.1	10.1	29.3	-19.5	16.5	45.9
Reactive Al ($\mu\text{g l}^{-1}$)	56	168	490	13	57	172
Labile Al ($\mu\text{g l}^{-1}$)	0	60	189	0	5	34
Strong acid ($\mu\text{equiv. l}^{-1}$)	-38	7.2	37	-175	-67	5.5
Weak acid ($\mu\text{equiv. l}^{-1}$)	49	129	253	72	116	185
Weak acid/TOC ($\mu\text{equiv. mg}^{-1}$)	9.5	15.8	55.7	8.2	15.6	33.2

* Sea-salt corrected values. Weak acid, $40 + 2.9 \text{ TOC} + 0.6 \text{ ILAL}$ ($r^2 = 0.812$) (Gulsvik); $51 + 3.5 \text{ TOC} + 0.6 \text{ ILAL}$ ($r^2 = 0.823$) (Nord-Trøndelag).

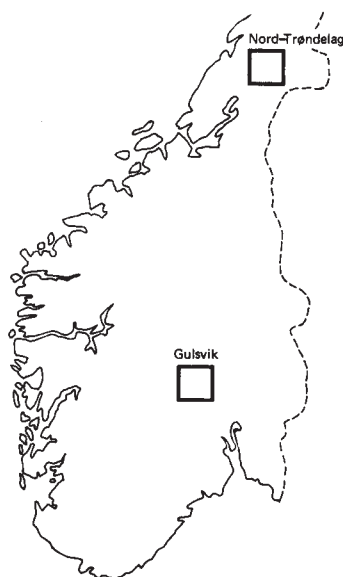


Fig. 1 Location of the Gulsvik and Nord-Trøndelag areas within Norway.

and TOC (Table 1), whereas mean base cations were lower in the lakes in southern Norway. The areas are separated by 550 km (Fig. 1). The area in southern Norway near Gulsvik receives precipitation with a lower pH (4.4 versus 4.9), higher concentrations of sulphate ($\text{SO}_4^{2-} = 54$ versus $12 \mu\text{equiv. l}^{-1}$) and lower contributions from sea salts (8.5 versus $68 \mu\text{equiv. l}^{-1} \text{Cl}^-$) than the more northern area (Nord-Trøndelag), where * denotes sea-salt corrected values.

Samples were collected from lake outlets in the early autumn of 1986 and were analysed at NIVA¹⁸. Cations were determined by flame atomic absorption spectrophotometry and anions by ion chromatography. TOC was measured by a carbon analyser. Methods for aluminium measurements are given in ref. 19. Strong and weak acids were determined by an automated coulometric Gran titration²⁰.

The contributions of TOC and sulphate to pH were evaluated in three ways, including comparisons of the measured values of strong and weak acidity in the two areas. First, measured strong acids were significantly higher in the lakes near Gulsvik. Strong acids were negative for nearly all of the lakes in Nord-Trøndelag ($\bar{x} = -67 \mu\text{equiv. l}^{-1}$) and positive for most of the lakes in southern Norway ($\bar{x} = 7.2 \mu\text{equiv. l}^{-1}$), whereas concentrations of weak acids and the amount of weak acid per milligram of TOC were similar for both areas (Table 1). Second, in both areas TOC and non-labile aluminium explained more than 81% of the variance in weak acids with nearly equivalent non-labile aluminium (ILAL) coefficients for the regressions.

The anion deficits ($\Sigma \text{Ca} + \text{Mg} + \text{Na} + \text{K} + \text{H} - \Sigma \text{HCO}_3 + \text{SO}_4 + \text{Cl} + \text{NO}_3$) in both sets of lakes were also similar and related to TOC by: $\text{DIFF} = 4.06 \text{ TOC} - 0.96$ ($r^2 = 0.551$) for all of the lakes. This relation is nearly identical to that in a recent survey of 1,000 lakes in Norway²¹, suggesting that the organic anion relation in the lakes sampled by this study is representative of large numbers and probably the population of low Ca lakes located in the granitic areas of Norway. Third, to assess the contribution of sulphate and organic anions to acidity, we can examine lakes with comparable TOC and base cation concentrations in the two areas (Table 2). Data from individual lakes show that the differences in the concentrations of strong acid (or alkalinity) between the two areas are a function of the difference in sea-salt corrected sulphate (SO_4^*) in the lakes (Table 2). The concentrations of strong acids can be evaluated from measured values and separately estimated from the concentrations of hydrogen ions and labile aluminium. The lakes near Gulsvik contained on average $43.6 \mu\text{equiv. l}^{-1}$ more SO_4^* than did lakes in Nord-Trøndelag. TOC contributes to strong acids in both areas, but the higher SO_4^* in the lakes in southern Norway explains the difference in strong acid concentrations between the areas and the lower pH values for the lakes in southern Norway.

The difference in concentrations of sulphate determines the difference in the pH of lakes having equivalent concentrations of base cations and TOC. This is demonstrated for lakes in both areas having comparable concentrations of TOC and base cations ($\Sigma \text{Ca}^* + \text{Mg}^* + \text{Na}^* + \text{K}^*$). The data are presented for individual lakes, because the comparisons must be made in relation to concentrations of base cations and ion balance considerations. Strong acidity results when the sum of strong acid anions exceeds base cations. For lakes G-21 and NT-101, which have low concentrations of $\text{Ca}^* + \text{Mg}^*$ (57 and $64 \mu\text{equiv. l}^{-1}$) or total base cations and high TOC (11.5 and 12.4 mg l^{-1}), the difference in the concentrations of measured strong acids is equal to the difference in excess sulphate (Table 2). At even lower concentrations of base cations (samples G-70 and NT-107) (33 and $34 \mu\text{equiv. l}^{-1}$) and TOC (5.6 and 8.8 mg l^{-1}), a similar difference is found in SO_4^* , which is only slightly higher than the difference in strong acids. At somewhat higher concentrations of base cations (54 and 54) but low TOC (2.7 and 3.8) (G-26 and NT-134), the total difference in strong acidity is also equivalent to the difference in SO_4^* . This is also true at higher $\text{Ca}^* + \text{Mg}^*$ (68 and 68) and TOC (8.7 and 9.4) (G-68 and NT-136). In addition, the concentration of strong acids is higher at constant SO_4^* with lower $\text{Ca}^* + \text{Mg}^*$, because in the Gulsvik area, strong acids = $7.97 - 0.84 \text{ Ca}^{2+} + 4.28 \text{ TOC} - 0.15$ weak acids + 0.44 SO_4^* ($r^2 = 0.864$; units are $\mu\text{equiv. l}^{-1}$ except TOC as mg l^{-1}). Thus, base cation concentrations regulate sensitivity, whereas SO_4^* and organic acids control pH. For all the lakes, strong acids = $-25.78 + 2.83 \text{ TOC} + 0.96 \text{ SO}_4^* - 0.98 \text{ Ca}^{2+}$ ($r^2 = 0.948$). We can conclude that with TOC and base cations held constant, the difference in strong acids between the areas is

Table 2 Comparisons of measured strong acidity and pH for lakes in two areas of Norway

Lake number	pH	Conductance (μS)	ALK ($\mu\text{equiv. l}^{-1}$)	TOC (mg l^{-1})	DIFF ($\mu\text{equiv. l}^{-1}$)	Strong acid ($\mu\text{equiv. l}^{-1}$)	Weak acid ($\mu\text{equiv. l}^{-1}$)	Cl (mg l^{-1})	$\text{NO}_3\text{-N}$ ($\mu\text{g l}^{-1}$)	RAL ($\mu\text{g l}^{-1}$)	LAL ($\mu\text{g l}^{-1}$)	$\text{Ca}^* + \text{Mg}^*$ ($\mu\text{equiv. l}^{-1}$)	SO_4^* ($\mu\text{equiv. l}^{-1}$)	$\text{Na}^* + \text{K}^*$ ($\mu\text{equiv. l}^{-1}$)
G-21	4.50	2.36	—	11.5	26	26	141	0.50	6	147	37	57	71	7
NT-101	5.62	3.04	18	12.4	56	-22	167	5.1	7	172	16	64	23	30
G-70	4.48	2.22	—	5.6	7	35	65	0.3	2	89	49	33	66	6
NT-107	5.32	2.24	5	8.8	37	1	130	3.8	6	107	17	34	19	21
G-68	4.65	2.32	—	8.7	35	25	139	0.6	10	204	98	68	63	7
NT-136	6.13	2.72	24	9.4	42	-30	147	4.8	2	97	5	68	15	9
G-26	5.37	1.36	—	2.7	8	-9	103	0.4	53	171	111	54	66	16
NT-134	6.27	2.28	27	3.8	20	-47	94	4.0	22	29	4	54	18	9
G-4	5.01	2.02	—	11.3	55	-6	157	0.6	1	213	55	108	77	14
NT-133	5.91	3.13	34	13.8	72	-54	185	5.2	2	114	7	103	20	15
NT-140	6.66	2.94	73	4.3	17	-61	79	4.8	—	16	1	95	21	10

G, Gulsvik; NT, Nord-Trøndelag; DIFF, $(\text{Ca} + \text{Mg} + \text{K} + \text{Na} + \text{H}) - (\text{HCO}_3 + \text{SO}_4 + \text{Cl} + \text{NO}_3)$; RAL, reactive aluminium; LAL, labile aluminium.

about equivalent to the difference in excess sulphate, and TOC contributes some strong acids in each area.

The organic contribution to acidity can be estimated from lakes NT-133 and 140 (Table 2). These lakes have similar $\text{Ca}^{2+} + \text{Mg}^{2+}$ (103 and 95) and equivalent SO_4^{2-} but differ in TOC by 9.5 mg l^{-1} (13.8 versus 4.3). The higher TOC lake (NT-133) has a lower pH (5.91 versus 6.66). Normalizing to equal base cation concentrations, the difference in alkalinity ($48 \mu\text{equiv. l}^{-1}$) corresponds closely to the difference in anion deficit between the lakes, thus explaining the difference in pH. The anion deficit (unmeasured organic anions) is $72 \mu\text{equiv. l}^{-1}$ compared with $17 \mu\text{equiv. l}^{-1}$ for the low TOC lake. This contribution is $5.5 \mu\text{equiv. mg}^{-1}$ TOC, close to the mean value for the sample of 1,000 lakes²¹. This pair of lakes demonstrates that organic anions can contribute significantly to lakewater acidity at approximately $4\text{--}5 \mu\text{equiv. mg}^{-1}$ TOC. While organic anions contribute to acidity, the high TOC lake maintains a pH well above levels toxic to brown trout.

One site in Nord-Trøndelag showed evidence of a pH depression due to the sea-salt effect^{22,23}. This was a small shallow pond draining a *Sphagnum* lawn. The moss effectively exchanged incoming Na^+ for H^+ ($\text{Na}^+ = -19.5 \mu\text{equiv. l}^{-1}$; $\text{Cl}^- = 155 \mu\text{equiv. l}^{-1}$) and the pH was 4.84. This bog lake was the only site that showed such an effect even though the contribution of sea-salt is relatively high in Nord-Trøndelag. The sea-salt effect typically produces only episodic pH depressions, and is restricted to coastal areas²²⁻²⁴. This effect did not contribute measurably to the acidity status of the lakes in either area.

We have shown that both organic acids and SO_4^{2-} contribute to strong acids. However, there is no evidence for a significant increase over time in the contributions of organic anions in lakes in southern Norway that would explain decreases in pH. It is therefore unlikely that organic anions have contributed to the recent significant decreases in lakewater alkalinity (acidification) and pH observed in southern Norway. The dark-water lakes in the Gulsvik area, although having sources of TOC in their watersheds, have anions dominated by SO_4^{2-} similar to clearwater acidic lakes in southernmost Norway²⁵. The organic acids make humic lakes more sensitive to acidic inputs from the atmosphere than clearwater lakes, for lakes with the same concentrations of base cations. This is also indicated by data from lakes in Nova Scotia¹². Organic acids originating from watersheds contribute to the acidity of the lakes, but not necessarily to the recent acidification of lakes. Rather, the higher values of strong acidity in southern Norway correlate with higher excess sulphate, because strong mineral acids are superimposed on organic contributions to acidity. The differences in strong acidity between lakes of comparable watershed factors and chemistry found by the present study are associated with the pattern of sulphate concentrations in lakes and the difference in precipitation acidity. This implies that the recent acidification of lakes in southern Norway, both clearwater lakes and those with TOC concentrations up to 15 mg l^{-1} , is a result of atmospheric sources of acids and not natural processes.

We acknowledge the support of the Norwegian Ministry of Environment and the Norwegian Institute for Water Research, sabbatical leave awards to D.F.B. and S.A.N. from Western Washington University and the University of Maine, respectively and a Royal Norwegian Council for Scientific and Industrial Research (NTNF) Senior Scientist Fellowship to D.F.B. Funded in part by the US Environmental Protection Agency (D.F.B.). We appreciate the assistance of Leif Lien in selecting lakes and sampling and comments from Eville Gorham and Richard F. Wright.

5. Krug, E. & Frink, C. R. *Science* **225**, 1427-1429 (1984).
6. Lazerte, B. D. & Dillon, P. J. *Can. J. Fish. Aquat. Sci.* **41**, 1664-1677 (1984).
7. Eschelman, K. N. & Hemond, H. F. *Wat. Resour. Res.* **21**, 1503-1510 (1985).
8. Clymo, R. S. *Phil. Trans. R. Soc. B305*, 487-499 (1984).
9. Gorham, E., Eisenreich, S. J., Ford, J. & Santelmann, M. V. in *Chemical Processes in Lakes* (ed. Stumm, W.) 339-363 (Wiley, New York, 1985).
10. Gorham, E. & Detenbeck, N. E. *J. Ecol.* **74**, 899-903 (1986).
11. Hemond, H. F. *Ecol. Monogr.* **50**, 507-526 (1980).
12. Gorham, E., Underwood, J. K., Martin, F. B. & Ogden, J. G. III *Nature* **324**, 451-453 (1986).
13. Battarbee, R. W., Flower, R. J., Stevenson, A. C. & Rippey, B. *Nature* **314**, 350-353 (1985).
14. Jones, V. J., Stevenson, A. C. & Battarbee, R. W. *Nature* **322**, 157-158 (1985).
15. Davis, R. B., Anderson, D. S. & Berge, F. *Nature* **316**, 436-438 (1985).
16. Davis, R. B., Norton, S. A., Hess, C. T. & Brakke, D. F. *Hydrobiologia* **103**, 113-123 (1983).
17. Charles, D. F. *Ecology* **66**, 994-1011 (1985).
18. Norton, S. A., Henriksen, A., Wathne, B. & Veidel, A. *Proc. Acidification and Water Pathways* 249-258 (Norwegian National Committee on Hydrology, Oslo, 1987).
19. Røgeberg, E. & Henriksen, A. *Vatten* **41**, 48-53 (1985).
20. Gjessing, E. & Røgeberg, E. *Acid Rain Res. Rep.* no. 12 (NIVA, Oslo, 1987).
21. Norwegian State Pollution Control Authority (SFT) *1,000 Lakes Survey 1986 Rep.* no. 187 (SFT, Oslo, 1987).
22. Skartveit, A. *Nord. Hydrol.* **1981**, 65-80 (1981).
23. Norton, S. A., Henriksen, A., Wright, R. F. & Brakke, D. F. in *Geochemistry and Monitoring in Representative Basins* (eds Moldan, B. & Paces, T.) 148-150 (Geological Survey, Prague, 1987).
24. Norton, S. A., Brakke, D. F. & Henriksen, A. *Environ. Sci. Technol.* (submitted).
25. Wright, R. F. & Henriksen, A. *Limnol. Oceanogr.* **23**, 487-498 (1978).

The age of the British chalk grassland

M. B. Bush* & J. R. Flenley†

* Department of Zoology, Ohio State University, 1753 Neil Avenue, Columbus, Ohio 43210, USA

† Department of Geography, University of Hull, Hull, North Humberside HU6 7RX, UK

Tansley¹ erected the hypothesis that the grasslands of the British chalk were the product of neolithic human forest clearance although he did not discount the possibility that in some areas the grasslands might have persisted throughout the forest period. Pigott and Walters² suggested a range of calcareous habitats where forests perhaps failed to form a closed canopy: where erosion was rapid, slopes steep, or soils too thin for trees to root. Despite these exceptions, most authors describe chalk grasslands as the product of neolithic clearance (for example, refs 3 and 4). The debate has been starved of evidence until now by the almost total lack of Flandrian botanical data from sites actually on the chalk. Here we summarize the results of a broad palaeoecological study (using pollen, and plant and animal microfossils) of a site in the heart of the Yorkshire Wolds which has yielded the first late-glacial and early Flandrian pollen record from the British chalk⁵. This deposit has provided evidence of early Flandrian chalk grasslands and of their persistence throughout the Pre-Boreal and Boreal periods.

Willow Garth is a 5-hectare carr lying in the Great Wold Valley (grid reference TA 126676) at an altitude of ~18 m Ordnance Datum. The woodland is flanked by arable fields which, when ploughed, reveal a patchwork of peaty and sandy deposits. To the north, Willow Garth is bounded by an underfit stream, the Gypsey Race⁶, which when it floods causes a pool to lie on the Willow Garth for some days. The carr has a more regular supply of water from at least five springs in the wood.

After an extensive stratigraphical survey (64 bore-holes and two ditch sections) the point of deepest organic accumulation (1.18 m) was identified (Fig. 1). This was a wet hollow ~20 × 30 m which supported a growth of *Iris pseudacorus*, *Filipendula ulmaria* and *Phalaris arundinacea*, shaded by *Salix fragilis*. The material for examination was collected from this point and was found to consist of 7 cm of undecomposed organic matter below which lay 41 cm of gyttja. From 48 cm to 118 cm there was a moss-peat, in which *Amblystegium riparium*, *Calliergon* spp., *Drepanocladus* spp. and *Scorpidium scorpioides* were the dominant taxa. Within this peat was a wood-rich band at ~76-82 cm. The basal contours of the peat were plotted and it was found that the sides of the hollow sloped gradually towards the

Received 9 March; accepted 3 July 1987.

1. Krug, E. C. & Frink, C. R. *Science* **221**, 520-525 (1983).
2. Krug, E. C., Isacson, P. J. & Frink, C. R. *J. air Poll. Control Ass.* **35**, 109-114 (1985).
3. Johnson, N. M., Likens, G. E., Feller, M. C. & Driscoll, C. T. *Science* **225**, 1424-1425 (1984).
4. Wright, R. F. *Science* **225**, 1426-1427 (1984).

Fig. 1 (right) The stratigraphy at Willow Garth. The core for analysis was collected at the intersection of transects CD and IJ.

Stratigraphic key

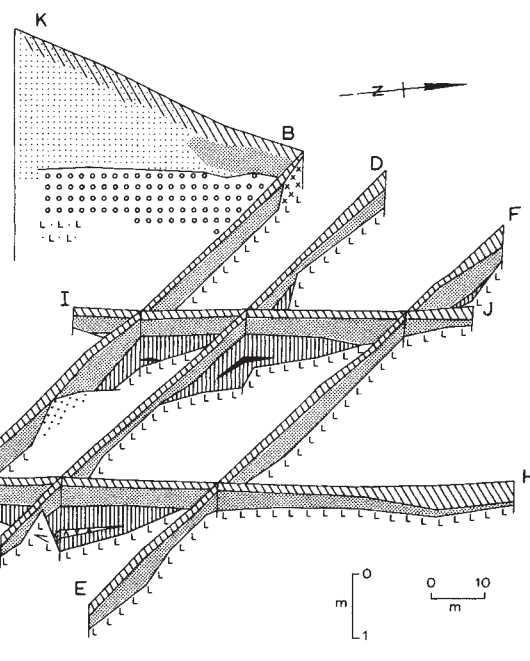
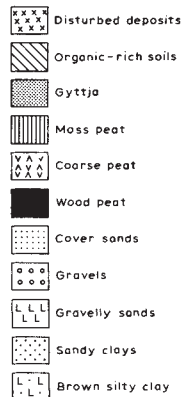
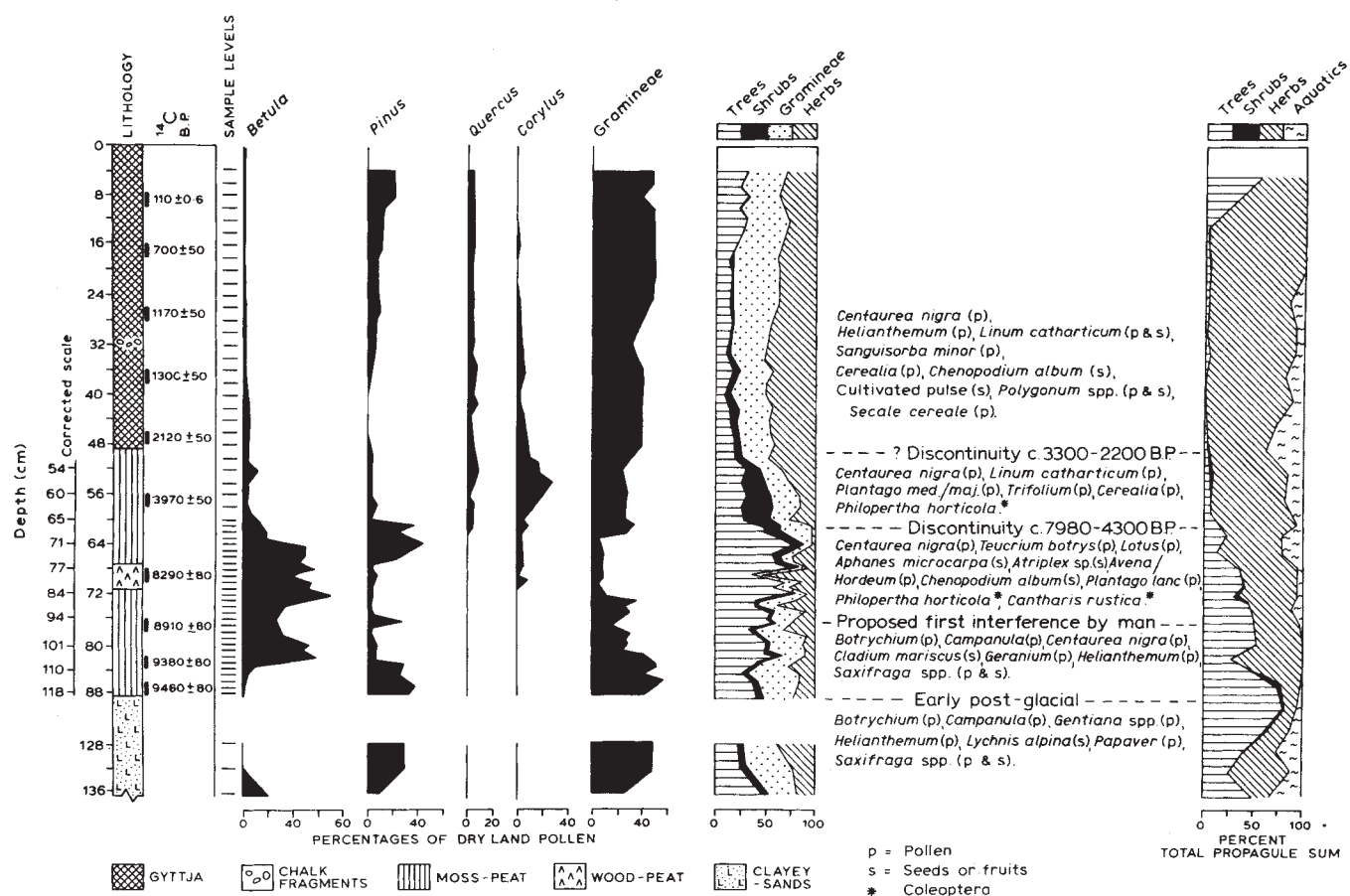


Fig. 2 (below) Summary pollen diagram from Willow Garth. Only selected taxa are shown.



centre. Sands with a low organic component lay beneath the peat (Fig. 1). Generally, these were coarse to gravelly sands, although those at the sample site were clay-rich (Fig. 2).

Using a Livingstone piston sampler (7 cm diameter) samples were collected for pollen analysis and radiocarbon dating. A monolith for macrofossil analysis was extracted from within 0.5 m of the pollen site. The pollen core suffered some compaction during collection but this could be corrected for, because a trial core, taken alongside using a D-section borer, showed the same total depth and stratigraphy but was not compressed.

All depths referred to in the text are 'corrected' depths.

Preparation of the pollen samples was by treatment with potassium hydroxide and acetolysis⁷, and by the use of 10-µm ultrasonic filtration⁸. Macrofossils were extracted as in ref. 9. Pollen counts had a range of 300-750 dry land pollen. Macrofossil counts were to a minimum of 200 higher plant propagules.

Ten radiocarbon dates were obtained (SRR-2665 to 2674, Fig. 2). The samples were examined, rootlets removed before dating, and only the 'organic carbon' used to establish ¹⁴C dates.

Old carbonate error does not appear to have been a problem as the wood-peat horizon at 76–82 cm offered a date consistent with those from the moss-peat and $\delta^{13}\text{C}$ values are normal. The base of the peat was dated at $9,460 \pm 80$ BP; however, the sandy material underlying the peat is, on the basis of pollen and macrofossils, estimated to have been deposited towards the end of the Windermere interstadial. The deposit seems to cover the period ~11,000 yr BP to ~7,980 yr BP (67 cm) when there was a standstill level. The record continued from ~4,300 yr BP through to the present day, although there may have been a further standstill level between ~3,300 and 2,200 yr BP.

In the basal sediments there is clear evidence of a cold-climate grassland flora with such species as *Gentiana verna*, *Diphasium alpinum*, *Lychnis alpina*, *Dryas octopetala*, *Saxifraga oppositifolia*, *Armeria maritima*, and *Plantago coronopus* (Fig. 2). Between ~9,380 and 9,300 yr BP *Betula* (probably *B. pubescens*) increased from ~10% to 58% of the dry land pollen count. However, this expansion was not maintained and *Betula* failed to exclude the open grassland taxa such as *Helianthemum*, *Campanula*, *Plantago coronopus* and *Botrychium lunaria*.

Betula declined to 24% (97 cm) and there was a peak of *Pinus* at 93 cm, immediately preceding the first regular occurrence of pollen and macrofossils of temperate arboreal taxa, such as *Corylus*, *Sambucus* and *Quercus*. However, these new arrivals failed to gain any kind of dominance and *Betula* showed a second peak of 68% at 84 cm.

Throughout this period there were high Gramineae levels, mostly 20 to 33% dry land pollen when, in most other British pollen diagrams, ~5% would be expected¹⁰. Grassland species had also persisted with *Plantago media* and/or *major*, *Campanula*, *Vicia*, *Lotus* and *Centaurea nigra* all present.

The failure of the temperate woodland taxa to dominate the pollen spectrum and the re-emergence of *Betula* suggests that the normal process of colonization had been halted. This does not seem to have been due to fire as there was no evidence of charcoal inwash, nor of peaks of fire-resistant taxa such as *Corylus*. The timber-rich layer of *Alnus*, *Salix* and *Quercus* wood suggests that there may have been some felling of the local forest. Although the two former species may have grown on the fen and collapsed into it, it is less likely that *Quercus* would have done so. The sudden occurrence of macrofossils of *Chenopodium album*, *Atriplex* cf. *patula*, *Aphanes arvensis*, and *Stellaria media* would support the suggestion that forest disturbance was taking place. Further evidence of disturbance was provided by the occurrence of insects characteristic of open grassland, such as *Philopertha horticola*, which was in every sample between 84 cm and 64 cm, *Cantharis rustica* and *Serrica brunea*. These insects all need dry grassland either in their immature or adult forms. The consistent presence of *P. horticola* is strongly suggestive of clearance (H. K. Kenward, personal communication). From this evidence it is suggested that man was disturbing the forest around Willow Garth from the beginning of the recovery of *Betula* (~8,900 yr BP). It is suggested that the resurgence of *Betula* in these circumstances would be as a secondary growth tree, or one which was able to flower despite cutting.

Cutting the trees would have had the effect of increasing young shoots for grazing wild cattle, could have increased the biomass of ungulates grazing the land by as much as 10-fold¹¹, and would have provided man with seeds from such plants as *Chenopodium album*, *Atriplex* cf. *patula* and *Avena*. Seven Gramineae pollen of >44 μm length were recovered from four samples between 93 and 88 cm (maximum size 52 μm with 10.5 μm annulus). These were compared with a range of reference material and identified as cf. *Avena*/*Hordeum* (presumably wild varieties).

Through this kind of management (forest clearance and grazing) the grasslands would have prospered as indicated by the presence of such taxa as *Teucrium botrys*, *Centaurea nigra* and *Sanguisorba minor*. The latter was found at a corresponding

level in a preliminary core, though not in the main pollen core.

No palaeoecological evidence is available from this site between ~7,980 and 4,300 yr BP, hence it is not possible to state whether these grasslands were the same as those which were evidenced above the standstill level. Open chalkland landscapes in early neolithic times (before 4,800 yr BP) have been described, for example at Kilham¹². Our results demonstrate the capacity of chalk grasslands, with the help of man, to survive the forest closure of the Pre-Boreal and Boreal periods, even in an area of gentle slopes. This leaves a period lasting from ~7,980 to 4,300 yr BP when the fate of the chalk grasslands has yet to be ascertained. Given the density of mesolithic and neolithic settlement in the Great Wold Valley¹³, it is likely that forest disturbance would have continued throughout this period. From 4,300 yr BP to the present the existence of grasslands is clear from Fig. 2. Hence, there is a strong probability that, in at least some areas of Britain, the chalk grasslands have been a continuous feature of the post-glacial landscape.

We thank Drs A. R. Hall and H. K. Kenward for their help in identifying macrofossil remains and the NERC radiocarbon committee for allocating funds for the ^{14}C dates.

Received 14 May; accepted 4 August 1987.

1. Tansley, A. G. *The Vegetation of the British Isles* (Cambridge University Press, 1939).
2. Pigott, C. D. & Walters, S. M. *J. Ecol.* **42**, 95–117 (1954).
3. Smith, C. J. *The Ecology of the English Chalk* (Academic, London, 1980).
4. Waton, P. V. thesis, Univ. Southampton, 1982).
5. Bush, M. B. thesis, Univ. Hull, 1986).
6. Lewin, S. *Univ. Hull. occ. Pap. Geogr.* **11**, 1–89 (1969).
7. Faegri, K. & Iversen, J. *A Textbook of Pollen Analysis* (1975).
8. Tomlinson, P. *Circaea* **2**, 139–141 (1984).
9. Kenward, H. K., Hall, A. R. & Jones, A. K. G. *Sci. Archaeol.* **22**, 3–15 (1984).
10. Godwin, H. *The History of the British Flora* 2nd edn (Cambridge University Press, 1975).
11. Mellars, P. *CBA Res. Rep.* **11**, 49–56 (1975).
12. Evans, J. G. & Dimbleby, G. W. *Proc. prehist. Soc.* **42**, 150–159 (1976).
13. Manby, T. G. *Yorks archaeol. J.* **47**, 23–61 (1975).

Upper Palaeolithic boomerang made of a mammoth tusk in south Poland

Paweł Valde-Nowak*, Adam Nadachowski† & Mieczysław Wolsan†

* Department of Archaeology of Little Poland, Institute of the History of Material Culture, Polish Academy of Sciences, Sławkowska 17, PL-31-016 Kraków, Poland

† Institute of Systematic and Experimental Zoology, Polish Academy of Sciences, Sławkowska 17, PL-31-016 Kraków, Poland

Excavations in a cave in the Oblazowa Rock in south Poland revealed an Upper Palaeolithic cultural level with an abundance of lithic and organic artefacts as well as palaeozoological material. This level yielded a fragment of mammoth tusk showing the characteristic traits of a complete boomerang. This is the earliest certain find of this type of weapon in the world. The same layer also contained an adult distal thumb phalanx which represents the earliest find of human skeletal material in Poland to date. Human food debris in the cultural layer suggests that the Upper Palaeolithic people inhabiting the cave hunted reindeer for the most part. The faunal complex of this cultural level is typical of the Upper Pleistocene steppe-tundra, while the artefacts belong within the complex of Central European cultures with backed points and have attributes which allow them to be grouped within the Pavlov culture dating from ~23,000 yr BP. The site might have been of importance in the course of migration of human groups with backed points from today's Moravia and southwestern Slovakia regions north-eastwards to the Don river basin.

Several definitions of the boomerang variously based on criteria such as material, purpose, aerodynamic qualities and shape occur in the vast literature on this weapon. Generally



Fig. 1 External (left) and internal (right) views of a fragment of mammoth tusk interpreted as a boomerang.

only returning specimens are recognized as boomerangs, others being excluded as 'killing sticks'^{1,2}. In this account we have adopted the definition presented in the latest monograph by H. Peter³ which is based on criteria of form (curvature and bilateral flattening), rather than 'returning ability', which may vary due to possible manufacturing faults, damage or deformation, and is impossible to verify in an archaeological specimen.

Today the boomerang is prevalent only in the Australian Aborigine culture, but in ancient times it was used on virtually all inhabited continents³. In Europe the Iron age site of Velsen in the Netherlands⁴ and the Mesolithic Brabrand-Sø in Jutland⁵ are known for discoveries of wooden boomerangs. At Stillfried in Lower Austria a mammoth bone fragment resembling a fragmented boomerang was found⁶ in an Upper Palaeolithic context but the exact chronology and culture are not certain⁷. The present discovery of a complete boomerang in Poland is thus the oldest definite find of this kind of weapon.

In summer 1985 the Institute of Systematic and Experimental Zoology, Polish Academy of Sciences, and the Department of Archaeology of Little Poland, Institute of the History of Material Culture, Polish Academy of Sciences, started excavations in a cave in the Oblazowa Rock in south Poland (49°25'48" N, 20°9'36" E). The cave is located in the lower part of the Oblazowa Rock (670 m above sea level) which, together with the Kramnica Rock (688 m above sea level), constitutes the inlet to the Bialka river gorge. The cave originated through decay and water erosion of a tectonic breccia in the red nodular limestone^{8,9}. A relatively small arched entrance to the cave, almost completely silted up before the excavations, is situated in the south wall of the rock about 6 m above the recent valley of the Bialka river. The greatest length of the cave exceeded 9 m while the breadth was over 4 m. About 12 m³ of the deposits has so far been excavated, resulting in exposure of the greater part of the entrance. However, to date the rock floor has not been detected.

Excavations have so far revealed nine sedimentary units containing abundant palaeozoological material and three Upper Palaeolithic cultural levels. The lowest of these levels has proved the most abundant in artefacts and contained part of a mammoth

tusk which exhibits the characteristics of a boomerang (Fig. 1). The span of the boomerang is about 70 cm, the greatest breadth 6 cm, and the greatest thickness about 1.5 cm. Its convex face is in fact the external tusk surface and exhibits some scratches which probably originated during the animal's life. On the other side it is almost flat as a result of polishing and shows several scratched lines without any decorative pattern, roughly parallel to the longer axis of the specimen. At one of its ends tiny transversal lines are visible on both sides. These lines are not decorative and presumably relate to the manufacture of the

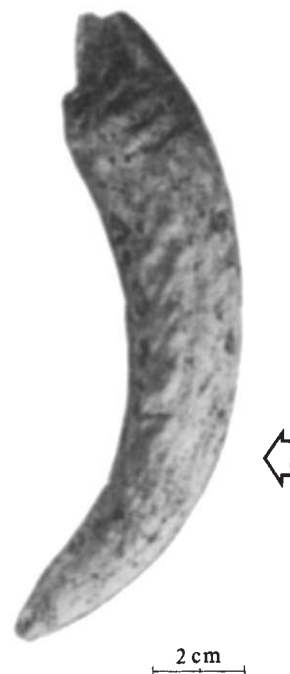


Fig. 2 Wedge made from the distal fragment of an antler of a large cervid, with incised decoration.

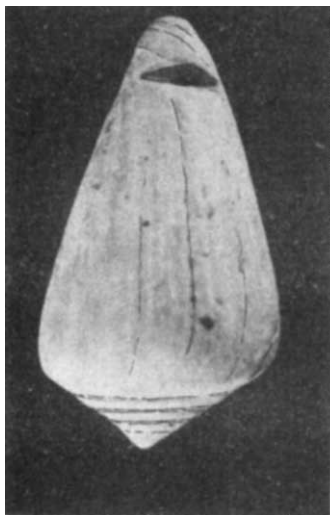


Fig. 3 Shell of a snail (*Conus* sp.) with an intentional aperture at the basal end. Assumed to have been used as bullroarer or pendant.



Fig. 4 Ventral view of the adult distal left thumb phalanx.

piece. Red colouring from a mineral substance was detected on some parts of the weapon, but there is no evidence to suggest that this is a result of painting.

Apart from the boomerang this lowest cultural level contained seven other artefacts made of organic materials. These included: a wedge made from the distal fragment of the antler from a large cervid, with a sharpened working end and decoration consisting of a few parallel arcuate lines (Fig. 2); a pendant or possible bullroarer made of a snail shell (*Conus* sp.) (Fig. 3); a necklace element made of the upper canine of an Arctic fox (*Alopex lagopus*); a perforator; and three bone beads. The assemblage of 30 stone artefacts comprise two flint cores and one of radiolarite, some blades, flakes and tools. Among the tools is an oblique truncated blade with undulating truncation (of Rgani type), a short end-scraper and a large tool of a cleaver type. The remains of a structure which consisted of several large granite and sandstone boulders up to 0.5 m in diameter were also found in this layer. Finally this cultural level also yielded the distal phalanx of an adult left thumb (Fig. 4), which represents the earliest find of human remains in Poland, as well as the bones and teeth of amphibians, reptiles, birds, and mammals, including Arctic lemming (*Dicrostonyx gulielmi*), Norway lemming (*Lemmus lemmus*), souslik (*Spermophilus* sp.), root vole (*Microtus oeconomus*), narrow-skulled vole (*Microtus gregalis*), snow vole (*Microtus nivalis*), pika (*Ochotona* sp.), hare (*Lepus* sp.), wolf (*Canis lupus*), Arctic fox (*Alopex lagopus*), stoat (*Mustela erminea*) and reindeer (*Rangifer tarandus*).

The faunal material from the cultural level containing the boomerang is not homogeneous in origin. The abundant fauna of small vertebrates undoubtedly derive from owl pellets, while some of the larger bones may represent food remains of other animal inhabitants of the cave as well as the bones of animals which died inside the cave. There is relatively little human food debris and this comprises heavily split reindeer bones for the most part, suggesting that the Palaeolithic people living in this cave hunted mainly this ungulate. The whole faunal complex indicates a cold period of the last glaciation when the extensive steppe-tundra covered the cave environs.

In general the artefacts of the lowest cultural level in question can be related to the complex of Central European cultures with backed points, showing traits of the Pavlov culture which is about 23,000 yr BP. The cultural characters of the inventory and the geographical position of the cave suggest that this site might

have been of importance in the course of migration of human groups with backed points from today's Moravia and south-western Slovakia north-eastwards to the Don river basin¹⁰.

We thank Dr E. Gleń-Haduch for anthropological expertise on the human skeletal element excavated, Dr M. Pawlikowski for mineralogical investigation of the red colour substance covering the boomerang and Drs J. Cook and A. W. Gentry for valuable comments.

Received 29 January; accepted 12 August 1987.

1. Thomas, J. *New Scientist* **99**, 838-843 (1983).
2. Bell, J. *New Scientist* **99**, 839 (1983).
3. Peter, H. *Wesen und Bedeutung des Bumerangs* (Braumüller, Vienna, 1986).
4. Hess, F. *Antiquity* **47**, 303-306 (1973).
5. Schwantes, G. *Deutschlands Urgeschichte* (Quelle and Meyer, Leipzig, 1934).
6. Kriegler, K. *Acta praehist.* **3/4**, 6-13 (1962).
7. Menghin, O. F. A. *Acta praehist.* **3/4**, 14-34 (1962).
8. Kowalski, K. *Jaskinie Polski* Vol. 3 (Państwowe Wydawnictwo Naukowe, Warsaw, 1954).
9. Birkenmajer, K. *Przewodnik geologiczny po pieninskim pasie skalnym* (Wydawnictwa Geologiczne, Warsaw, 1979).
10. Kozłowski, J. K. & Kozłowski, S. K. *Pradzieje Europy od XL do IV tysiąclecia p.n.e.* (Państwowe Wydawnictwo Naukowe, Warsaw, 1975).

Endstopped neurons in the visual cortex as a substrate for calculating curvature

Allan Dobbins, Steven W. Zucker & Max S. Cynader*

Computer Vision and Robotics Laboratory, McGill Research Centre for Intelligent Machines, McGill University, 3480 University Street, Montréal, Québec H3A 2A7, Canada

Neurons in the visual cortex typically respond selectively to the orientation, and velocity and direction of movement, of moving-bar stimuli. These responses are generally thought to provide information about the orientation and position of lines and edges in the visual field. Some cells are also endstopped, that is selective for bars of specific lengths. Hubel and Wiesel first observed that endstopped hypercomplex cells¹ could respond to curved stimuli and suggested they might be involved in detection of curvature, but the exact relationship between endstopping and curvature has never been determined. We present here a mathematical model relating endstopping to curvature in which the difference in response of two simple cells gives rise to endstopping and varies in proportion to curvature. We also provide physiological evidence that endstopped cells in area 17 of the cat visual cortex are selective for curvature, whereas non-endstopped cells are not, and that some are selective for the sign of curvature. The prevailing view of edge and curve determination is that orientations are selected locally by the class of simple cortical cells³ and then integrated to form global curves. We have developed a computational theory of orientation selection^{4,5} which shows that measurements of orientation obtained by simple cells are not sufficient because there will be strong, incorrect responses from cells whose receptive fields (RFs) span distinct curves (Fig. 1). If estimates of curvature are available, however, these inappropriate responses can be eliminated. Curvature provides the key to structuring the network that underlies our theory and distinguishes it from previous lateral inhibition schemes^{6,7}.

The defining characteristic of endstopped cells is the presence of inhibitory receptive field (RF) end zones that 'stop' the response of the cell to stimuli which are long enough to intrude into the end zones. Two alternative models of endstopping have been proposed^{1,2}: in one, the inhibitory 'endzones' derive from two or more independent cortical afferents with offset RFs; in the other, they derive from a single inhibitory input with a long RF. Both are compatible with the requirement that the endzones

* Departments of Physiology, Biophysics and Psychology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada.

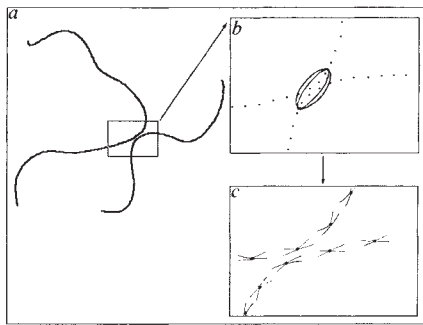


Fig. 1 Illustration of the use of curvature information in orientation selection. *a*, Two distinct curves closely approach each other so that the tangent to a point on one of the curves is colinear with the tangent to a nearby point on the other curve. *b*, This region is schematically expanded. In simulations, a simple cell-like operator centred between the curves and spanning both is strongly activated. This leads to an incorrect representation of the two distinct arcs as a single, unresolved curve. *c*, Each cone-like segment represents bounds on the estimate of curvature at a point. If such estimates were available, for example from the model that we propose, then the incorrect simple cell responses linking the two curves could be eliminated. A network for accomplishing this is described in detail in Parent and Zucker¹⁵.

share the orientation preference of the excitatory centre^{1,8}. Current evidence favours inhibition by a single large RF cell⁹ for reversible inactivation of layer VI, which contains the long RF cells¹⁰, eliminated or reduced endstopping in layer IV and the superficial layers; there is also anatomical evidence for a layer VI to IV connection^{11,12}.

With this background, observe that curvature can be defined as the rate of change of the tangent direction along a curve. For

straight lines there is no change, whereas for circles the change is constant. Therefore, curvature varies directly with deviation from straightness over a local neighbourhood through which the curve passes.

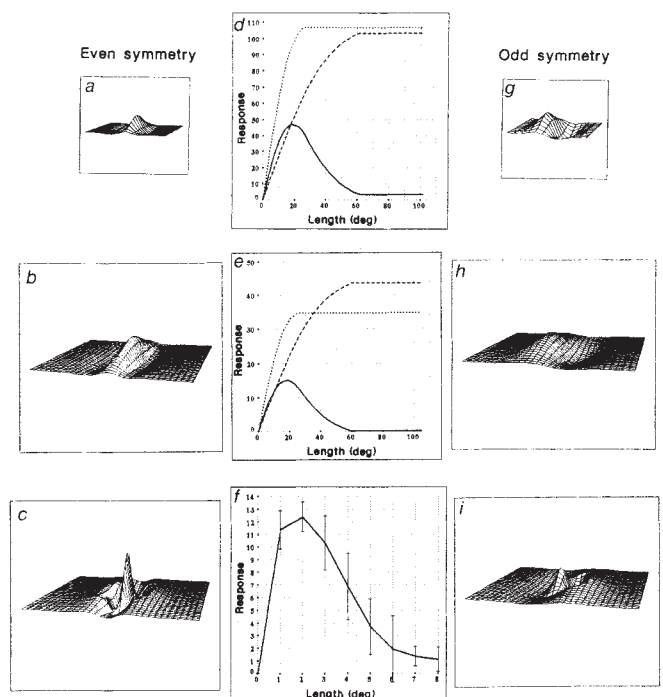
In our model for estimating curvature, two simple cells estimate the straightness of a contour, one over a small neighbourhood and the other over a larger one, and the difference in their responses, which converge upon an endstopped simple cell (ES), varies with the deviation from straightness, that is, the curvature. Consider two similarly positioned and oriented simple cells with RFs of different size, termed the small and large simple cells. An endstopped simple cell receives excitation from the small cell and inhibition from the large cell, thereby encoding both orientation and curvature information. In symbols, let R_S denote the response of the small simple cell, R_L the response from the large simple cell to a particular contrast pattern at the same position and orientation, $\phi(\cdot)$ a rectifying or clipping function equal to its argument when positive (when excitation sufficiently exceeds inhibition) and zero otherwise, then the end-inhibited simple cell response is given by

$$R_{ES} = \phi(c_S \cdot \phi(R_S) - c_L \cdot \phi(R_L))$$

where c_S and c_L are positive constants that normalize the area difference between the receptive fields (Fig. 4). The non-linear clipping in the model increases the stimulus specificity by preventing a negative value for R_L from being flipped into a positive contribution to R_{ES} , the response of the ES cell. $\phi(\cdot)$ models the inability of spike trains to code negative numbers against a background of low spontaneous activity such as that found in area 17. The simple RFs can have either even or odd symmetry, that is, can be either symmetrical or anti-symmetrical about the central axis and Fig. 2 shows ES models composed of differences of each type.

Fig. 2 The endstopped simple (ES) model. Binocular and temporal properties, and their non-linearities, are not incorporated, allowing the retinogeniculo-cortical pathway to be collapsed into a spatial receptive field, so the response of the cell can be modelled as a two-dimensional spatial convolution. Each simple RF may be modelled as a Gabor function or a difference of Gaussians (DOG). *a, b, c*, (Left column) Representation of the RF profiles of two even-symmetric simple cells and their linear combination. Each simple RF has a Gaussian weighting along the long axis and a DOG in the perpendicular direction. For these two simple RFs, the width ratio of the Gaussians comprising the DOG (σ_{x2}/σ_{x1}) is 2.5. The aspect ratios (σ_y/σ_{x1}) are 4 and 5 for the small and large RFs respectively. The smaller RF is 25 units long and the larger is 61. A linear combination of the two is also shown with weight constants $C_S = 3$ and $C_L = 1$. This graphic representation corresponds to the ES model only when the results of both simple convolutions are positive. (These parameter values are also used in Figure 3*d*.) *d, e, f* (Middle column), Length responses of even- and odd-symmetry ES models and of an end-stopped neuron respectively. *d*, The response of even-symmetry ES cell model to optimally-oriented bars of different length centered on the RF. The ES-cell model (solid curve) and the component responses of the two simple cells ($C_S \cdot \phi(R_S)$ and $C_L \cdot \phi(R_L)$, the dotted and dashed curves, respectively) were computed for bars 3 units wide with 50 lengths ranging from 2 to 100 units. *e*, The length-response of the odd-symmetry ES cell model (solid, dotted and dashed curves as in *d*). *f*, The actual response to moving bar stimuli of varying length obtained from an end-stopped neuron. *g, h, i* (Right column), Representation of RF profiles of two odd-symmetry simple cells and their linear combination.

Methods. Surgery was performed under sodium pentothal anaesthesia with marcaine (a long-lasting local anaesthetic) applied around the surgical wounds. During the experiment, paralysis was maintained with continuous infusion of flaxedil ($10 \text{ mg kg}^{-1} \text{ h}^{-1}$) with anaesthesia by a respirated mixture of 70% $\text{N}_2\text{O}/30\% \text{ O}_2$ supplemented with intravenously infused nembutal ($1 \text{ mg kg}^{-1} \text{ h}^{-1}$). The animal was also infused with 5% dextrose in lactated Ringer solution (2 ml h^{-1}). Rectal temperature, end-tidal CO_2 and electrocardiogram were continuously monitored. Neuronal responses were assessed quantitatively by measuring: length tuning using optimally-oriented bar stimuli moving in the preferred direction, line-weighting functions (bars of the preferred orientation flashed throughout the central RF), and curvature-tuning with semi-circular arcs spanning a wide range of radii, moved in the preferred direction and positioned so that the tangent of the curve at mid-arc passes over the central RF parallel to the preferred orientation of the RF. For a more detailed description of our surgical and recording procedures see Ref. 13. *g, h, i*, The two odd-symmetry simple RFs are described by Gabor functions (a 2-D Gaussian multiplied by a sinusoid). For the small and large simple RFs (*g, h*), the aspect ratio (σ_y/σ_x) is 3 and the period ratio (λ/σ_x) is 2, where λ is the period of the sine function. *i*, Shows a linear combination of the two ($C_S = 2.5$ and $C_L = 1$). Again, the non-linear components of the model are omitted.



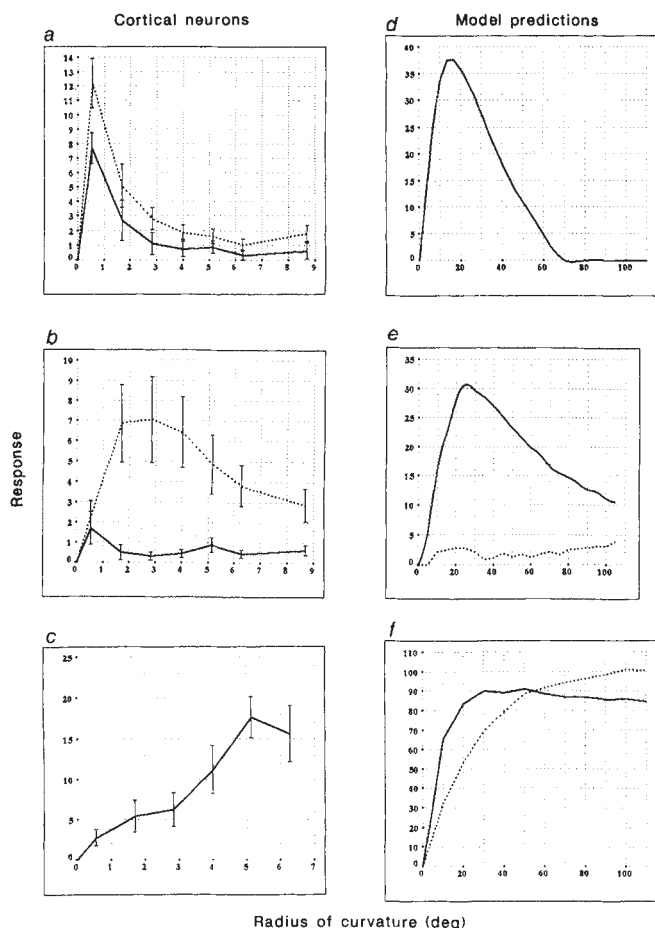


Fig. 3 Experimental (*d, e, f*) and theoretical (*a, b, c*) curvature-response data. Solid and dotted curves correspond to the two signs of curvature in all parts of the figure except *f*. *a, b, c* (Left column), Curvature-response of three area 17 neurons evaluated with semi-circular white arcs projected onto a darker screen. The tangent at mid-arc was parallel to the preferred orientation of the RF and moved in the preferred direction. The curves are based on sample means and standard errors from 16 randomly-interleaved presentations. A cubic spline was used to interpolate smoothly between the points in *a-e*. *a*, The curvature response of the endstopped neuron for which the length response is shown in Fig. 2*f*. The response does not vary with sign of curvature. *b*, Curvature response of an endstopped neuron. *c*, Curvature response of a nonendstopped neurone. *d-f* (Right column), Curvature response of even- and odd-symmetry instances of the model and the component responses of the even model produced by convolving the ES model with semi-circular arcs 3 units wide with different radii, to illustrate the sensitivity of the ES model to curvature. The curves are positioned so that the tangent to the curve at the point on the RF centre is parallel to the long axis of the RF. In each case the model was convolved with 20 curves varying in radius from 5 to 100 units. *d*, Curvature response of the even-symmetry ES model with the parameter values given in Fig. 2. *e*, Curvature response of the odd-symmetry model of Fig. 2 for the two signs of curvature, showing how different combinations of small and large cells give rise to different curvature selectivity and in particular, providing a suggestion for how different responses to the two signs of curvature could arise. *f*, Curvature response of the two simple components of the ES model in *d*, demonstrating how curvature selectivity arises. The non-endstopped neuron (*c*) and non-endstopped simulations (*f*) have a much broader band pass than the endstopped neurons and simulations.

Responses as a function of length for the even- and odd-symmetry ES models and their simple RF components are shown in Fig. 2*d, e*: a typical length response curve obtained in one of our neurophysiological experiments is shown for comparison (Fig. 2*f*). By varying the size of the two receptive fields, within

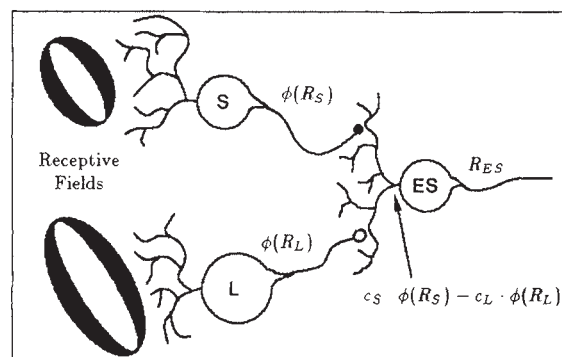


Fig. 4 A schematic model of the endstopped simple circuit. Neurons classified as simple cells with small (S) and large (L) receptive fields provide excitation and inhibition, respectively, to the endstopped simple cell (ES) upon which they converge.

physiological limits, and the weighting constants, it is possible to replicate any of the length-response data found in the literature while preserving position and orientation tuning. Although we shall not consider them further here, various combinations of weights and sizes are necessary to deal with the contrast and width variations that arise in natural images¹⁴.

By analogy with a length-response curve, a curvature-response curve can be obtained by evaluating the model over a series of semi-circular arcs with increasing radius (Fig. 3). Two possibilities arise, one in which the model is composed of even-symmetry simple cells, and one in which they have odd symmetry. Consider the even-symmetry model first. For a stimulus with a low radius of curvature, the curve passes through the excitatory region of the receptive field of the small simple cell but does not pass through enough of the excitatory region of the receptive field of the large simple cell to inhibit it effectively. As the radius of curvature increases, the stimulus becomes increasingly effective in driving the large receptive field and the inhibition increases. Figure 3*d, f* shows the curvature response for the ES model and its even-symmetry components. For the odd-symmetry ES model, the curvature-response depends on the sign of curvature (Fig. 3*e*). For one sign of curvature, a curved stimulus excites the small RF but curves into the antagonistic region of the large RF, reducing its response and hence its inhibitory contribution to the ES cell. A similarly curved stimulus of the opposite sign passes through only the excitatory region of the large RF, allowing effective inhibition. For both signs of curvature, as the radius of curvature increases, inhibition becomes increasingly effective.

These computational results encouraged us to seek evidence for curvature-selective neurons in area 17 in the visual cortex of the cat. We will describe a cell as curvature-selective, if it responds strongly over only a limited range of curvatures when stimulated with a continuous arc falling within a range of positions and orientations at the centre of the RF. The curvature response of both endstopped and non-endstopped cells was examined to determine if curvature-selectivity is systematically related to endstopping. Cells were qualitatively characterized using a hand projector and then examined quantitatively using computer-controlled optics.

All of the fifteen endstopped neurons, both simple and complex, examined carefully so far exhibit the predicted peaked curvature response (Fig. 3*a, b*) whereas non-endstopped neurons exhibit a different, non-peaked response (Fig. 3*c*). The neuron of Fig. 3*a* displays similar curvature tuning for both signs of curvature, qualitatively resembling the even-symmetry model, whereas the one in Fig. 3*b* has virtually no response for one direction of curvature, similar to the odd-symmetry model. More than a third of the neurons exhibited significantly stronger responses to one direction of curvature than to the other, the cell in Fig. 3*b* being one of the most striking examples. Although

most endstopped cells are directionally selective, we tested curvature response in both the preferred and null (opposite) directions of motion for some of the less directionally sensitive units. The results indicate similar curvature selectivity for both directions of motion.

In comparing the single unit results to the predictions of the ES model, note that layer VI contains complex RF types as well as simple, and they may also contribute to the generation of end inhibition and curvature-selective responses. Even so, an asymmetrical contribution from the large RF cell, whether of spatial or temporal origin, is required to generate responses selective for sign of curvature. Because experiments to determine the basis of this selectivity have not yet been performed, inhibition from a large, odd-symmetry inhibitory RF (Figs 2 and 3) must be considered as only one of the possible mechanisms.

Obtaining straightness (tangent) and curvature estimates from a difference of two first-stage responses has two advantages: anatomically, only vertical connections within a cortical column are required, and numerically, there is greater noise immunity in a first-difference operation than in repeated-difference operations.

Endstopping is usually thought to be related to length or end-point detection. We provide evidence for an alternative,

that orientation-selective end inhibition allows the calculation of coarse-curvature estimates. Such estimates facilitate the process of orientation selection, and could be important in shape representation. As there are far more curves than end points in natural images, this proposal has a certain ecological plausibility.

We thank L. Iverson for programming support. Research was supported by grants from DREA, MRC and NSERC (S.W.Z., M.S.C.). S.W.Z. and M.S.C. are Senior Fellows of the Canadian Institute of Advanced Research.

Received 26 May; accepted 7 August 1987.

- Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **28**, 229-289 (1965).
- Dreher, B. *Invest. Ophthalmol.* **11**, 355-356 (1972).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106-154 (1962).
- Zucker, S. W. *Comp. Vis. Gr. Im. Proc.* **74**-103 (1985).
- Zucker, S. W. *Behav. Res. Meth., Instr. Comp.* **18**, 608-617 (1986).
- Blakemore, C., Carpenter, R. & Georgeson, M. *Nature* **228**, 37-39 (1970).
- Carpenter, R. & Blakemore, C. *Expl Brain Res.* **18**, 287-303 (1973).
- Orban, G. A., Kato, H. & Bishop, P. O. *J. Neurophysiol.* **42**, 833-849 (1979).
- Bolz, J. & Gilbert, C. D. *Nature* **320**, 362-365 (1986).
- Gilbert, C. D. *J. Physiol.* **268**, 391-421 (1977).
- McGuire, B. A., Hornung, J.-P., Gilbert, C. D. & Wiesel, T. N. *J. Neurosci.* **4**, 3021-3033 (1984).
- Wiesel, T. N. & Gilbert, C. D. *Quart. J. exp. Physiol.* **68**, 525-543 (1983).
- Baker, C. L. & Cynader, M. S. *J. Neurophysiol.* **55**, 1136-1152 (1986).
- Dobbins, A. *thesis*, McGill University, 1987.
- Parent, P. & Zucker, S. W. *IEEE Trans. PAMI* (in the press).

Kindling stimulation induces *c-fos* protein(s) in granule cells of the rat dentate gyrus

M. Dragunow & H. A. Robertson

Department of Pharmacology, Faculty of Medicine,
Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

Alterations in neuronal gene expression have been proposed to account for permanent changes in brain function such as learning and memory¹⁻⁴. In particular, it has been suggested that proto-oncogenes such as *c-fos* (ref. 5) may be rapidly induced in conditions that lead to neuronal plasticity and evoke permanent changes in the expression of effector genes¹. Concentrations of the *c-fos* proto-oncogene increase rapidly following depolarization-induced calcium influx in non-dividing neuronally differentiated PC 12 cells^{2,3}. Recently, the presence and induction of *c-fos* in the adult brain and spinal cord has been observed⁶⁻⁸. Here we report that electrically-induced seizure activity, which leads to a permanent increase in the response of the brain to future seizures (kindling⁹), rapidly and transiently increases *c-fos* protein-like immunoreactivity in the nuclei of granule cells in the rat dentate gyrus. These results suggest that *c-fos* protein is present within the nuclei of adult mammalian neurons, and could be involved in plastic changes in the nervous system associated with seizure activity.

To study the effects of kindling stimulation on the production of *c-fos* protein(s), 40 adult male Sprague-Dawley rats were anaesthetized with barbiturate and bipolar stainless steel electrodes were implanted unilaterally into the dorsal hippocampus as previously described¹⁰. At least one week later, rats were electrically stimulated with high frequency trains (kindling stimuli: 500 μ A, 100 Hz, 0.5 ms/half-wave diphasic pulses for 2 s) which produced brief partial hippocampal seizures, comprising an initial burst of high frequency and amplitude spiking lasting from 10-20 s, followed by a post-ictal electroencephalogram silence that was terminated by a second bout of afterdischarge lasting from 10-30 s. At various times after the spontaneous termination of these partial hippocampal seizures, rats were anaesthetized with barbiturate, and perfused and fixed for *c-fos* protein immunohistochemistry as previously described⁶.

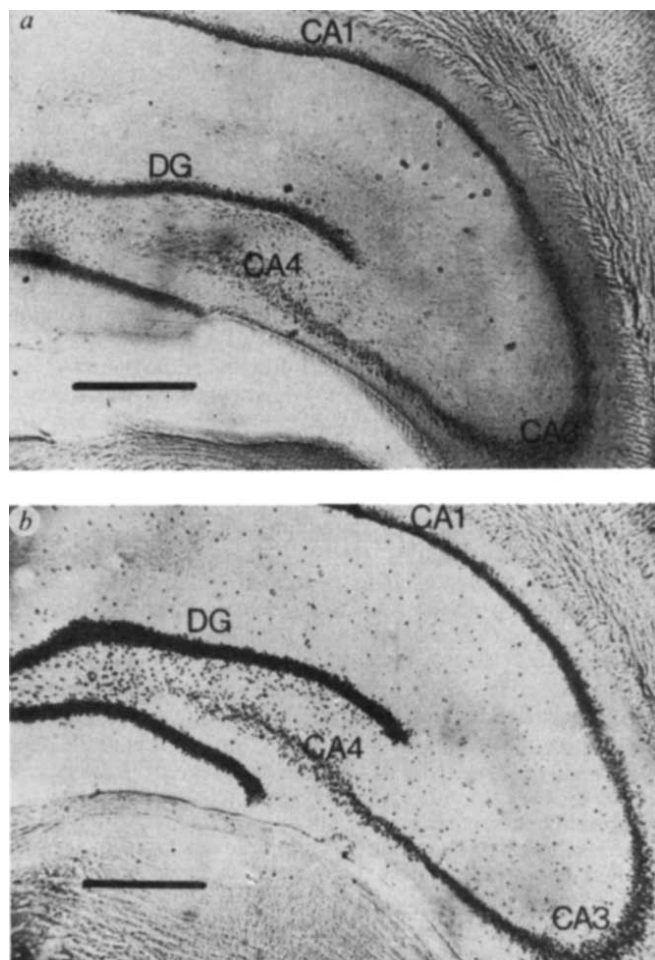


Fig. 1 *a*, Photomicrograph of rat hippocampal formation (areas CA1, CA3, and CA4) and dentate gyrus (DG) from an electrode implanted unstimulated rat. *b*, Photomicrograph of rat hippocampal formation and dentate gyrus fixed 30 min following a hippocampal seizure. Note the intense immunoreactivity in dentate granule cells and the increase in immunostaining in the pyramidal cells of areas CA1, CA3 and CA4. Antiserum was diluted 1 in 1,000 (Scale bar, 0.5 mm).

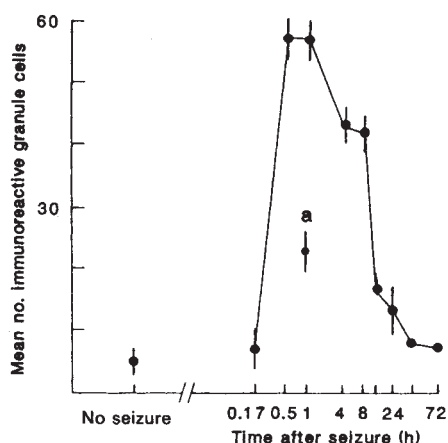


Fig. 2 Graph showing the average number of granule cells having positive *c-fos* protein immunoreactivity in an area of the dentate gyrus bounded by a 100 μm square. Values (mean $\pm$ standard deviation) were taken before, and at various times after, seizure activity at coordinates anterior/posterior -4.3 from bregma, using the atlas of Paxinos and Watson¹⁷. Point (a) denotes rats injected intraperitoneally with 40 mg per kg carbamazepine 30 min prior to seizure. Note the great reduction in immunostaining 1 h post-stimulation ($T = 3.4$, 9 d.f., $P < 0.05$).

Positive *c-fos* protein-like immunoreactivity was found within the nucleus of nerve cells (but not in glia) throughout the brain in a pattern similar to that found previously^{6,7} (Fig. 1a). Neurons in the cerebral cortex, hippocampus, dentate gyrus, amygdala, caudate and piriform cortices showed positive immunostaining. Preadsorbing the antibody with the *c-fos* peptide¹¹ abolished the immunostaining. This result, together with the observed nuclear location of the immunostaining where *c-fos* proteins are known to accumulate^{12,13}, suggests that the immunostaining is to an endogenous *c-fos* protein-like molecule(s).

Rats that experienced seizure activity after electrical kindling stimulation showed a massive increase in *c-fos* protein-like immunoreactivity bilaterally in granule cells of the dentate gyrus (Fig. 1b and 2). This increase in immunostaining was rapid in onset, being present at 30 but not 10 min after seizure, and transient; the number of cells showing the response was almost back to baseline after 24 hours (Fig. 2). Immunostaining for *c-fos* protein was also increased in fields CA1, CA3 and CA4 of the hippocampus with a slower time-course (Fig. 1b), and in the piriform cortex. In rats that were electrically stimulated but treated with carbamazepine before stimulation to prevent seizures, a small increase in immunostaining, confined to the stimulated side, was observed one hour later (Fig. 2).

This seizure-related increase in *c-fos* protein immunoreactivity could arise from a number of possible factors, including the increased production of *c-fos* protein and interference with *c-fos* protein degradation. The latent period of onset of the effect corresponds to the onset of *de novo* synthesis of *c-fos* protein in culture systems stimulated by growth factors¹⁴, suggesting that the increase is due to *de novo* synthesis of *c-fos* protein. This interpretation is supported by the work of Curran and his collaborators⁷, who showed by Northern blot analysis that metrazol seizures increase the amount of *c-fos* messenger RNA.

These results demonstrate first that the *c-fos* protein(s) is present in the nuclei of adult rat neurons, which confirms previous results⁶⁻⁸, and second, that activation of nerve cells with seizure activity producing kindling¹⁰ greatly increases the production of the *c-fos* protein in granule cells of the dentate gyrus. The mechanism by which kindling-related seizure activity induces *c-fos* in granule cells is unclear. Calcium is known to accumulate in nerve cells during seizures¹⁵, however, and induction of *c-fos* by nicotine and potassium in PC12 cells is calcium-dependent^{2,3}. The molecular layer of the dentate gyrus has one

of the highest densities of calcium antagonist binding sites in the rat brain¹⁶ and this could account for the massive and rapid increase in *c-fos* in the dentate following seizures. The increase in *c-fos* is one of the earliest genetic events to follow neuronal activation³. It could fulfill a critical early function in coupling relatively brief electrical events to other long-term transcription-dependent events that maintain long-term changes in the nervous system, such as those observed in kindling.

We are grateful to Tom Curran for the *c-fos* antibody. This work was supported by grants from the New Zealand Neurological Foundation and the Canadian Medical Research Council.

Received 18 June; accepted 25 August 1987.

1. Goebel, P., Castellucci, V. F., Schacher, S. & Kandel, E. R. *Nature* **322**, 419-422 (1986).
2. Greenberg, M. E., Ziff, E. B. & Greene, L. A. *Science* **234**, 80-83 (1986).
3. Morgan, J. I. & Curran, T. *Nature* **322**, 552-555 (1986).
4. Montarolo, P. G. *et al. Science* **234**, 1429-1434 (1986).
5. Bishop, J. M. A. *Rev. Biochem.* **53**, 301-354 (1983).
6. Dragunow, M., Peterson, M. R. & Robertson, H. A. *Eur. J. Pharmacol.* **135**, 113-114 (1987).
7. Morgan, J. I., Cohen, D. R., Hempstead, J. L. & Curran, T. *Science* **237**, 192-197 (1987).
8. Hunt, S. P., Pini, A. & Evan, G. *Nature* **328**, 632-635 (1987).
9. Goddard, G. V. *Nature* **214**, 1020-1021 (1967).
10. Dragunow, M. *Neurosci. Lett.* **74**, 320-324 (1987).
11. Curran, T., Van Beveren, C., Ling, N. & Verma, I. M. *Molec. cell. Biol.* **5**, 167-172 (1985).
12. Adamson, E. D., Meek, J. & Edwards, S. A. *EMBO J.* **4**, 941-947 (1985).
13. Sambucetti, L. C. & Curran, T. *Science* **234**, 1417-1419 (1986).
14. Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. M. *Nature* **312**, 711-716 (1984).
15. Wong, R. K. S. & Prince, D. A. *Brain Res.* **234**, 385-390 (1978).
16. Cortes, R., Supavilai, P., Karobath, M. & Palacios, J. M. *J. neural Trans.* **60**, 169-197 (1984).
17. Paxinos, G. & Watson, C. in *The Rat Brain in Stereotaxic Coordinates* (Academic, Sydney, 1982).

EDRF coordinates the behaviour of vascular resistance vessels

T. M. Griffith, D. H. Edwards*, R. Ll. Davies, T. J. Harrison & K. T. Evans

Departments of Diagnostic Radiology & *Cardiology, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, UK

Constriction of vascular smooth muscle in response to the stimulus of raised intravascular pressure—the myogenic response^{1,2}—represents a positive feedback mechanism which, if unopposed, could theoretically lead to instability in the intact circulation^{3,4}. Dilation in response to increased intraluminal flow would provide an opposing feedback mechanism which could confer overall stability⁴. Flow-dependent dilation in conduit vessels⁵⁻⁷ is mediated by endothelium-derived relaxing factor (EDRF)⁸⁻¹⁴, but the relationship between flow and EDRF activity has not been studied in resistance vessels *in situ*. We here demonstrate that EDRF can coordinate the aggregate hydrodynamic properties of an intact network. Under control conditions, EDRF maintains a fourth-power relationship between diameter and flow so that the pressure gradient in each vessel asymptotically approaches a constant value at high flow rates. Basal EDRF release may also maintain a similar spatial distribution of flow at different flow rates, even under conditions of moderate pharmacological constriction.

Vascular endothelium releases EDRF when stimulated by a variety of pharmacological agents¹⁵. The interaction of basally released EDRF¹⁶⁻¹⁹ with myogenic responses in an intact vascular network might however be one of its major physiological functions. We have used a novel microangiographic technique²⁰ to investigate the influence of basal EDRF activity on diameter-flow relationships in an isolated, intact rabbit ear preparation (which has the advantage of being two-dimensional and relatively free of complicating metabolic influences) perfused under conditions of controlled, steady flow with physiological buffer. The method allows measurement of vessel calibre to the third generation of branch arteries ($\sim 70 \mu\text{m}$ in diameter). EDRF has

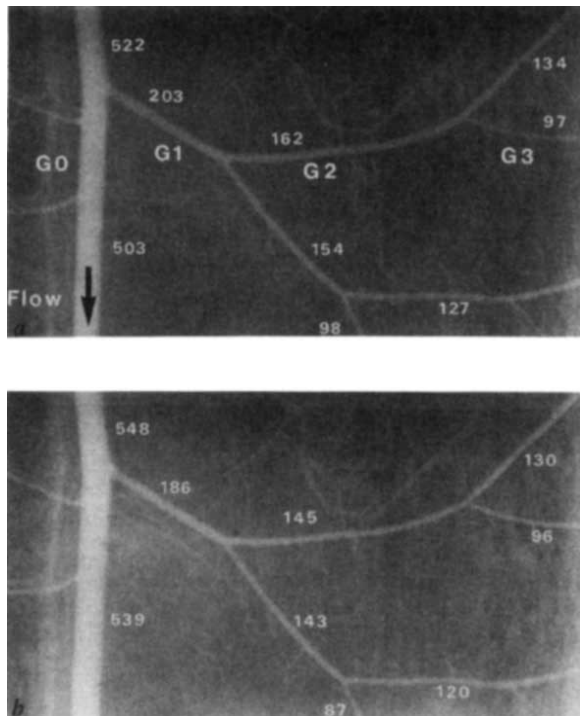


Fig. 1 Representative microangiograms of rabbit ear perfused at constant flow (2 ml min^{-1}) without (*a*) and with (*b*) $1 \mu\text{M}$ haemoglobin. Haemoglobin increased the pressure drop across the preparation from 35 mmHg to 50 mmHg , dilated the central ear artery (G0), and constricted the first, second and third generations of branch arteries (G1, G2 and G3). Vessel diameters (in μm) are shown.

Methods. Ears were removed from 2.5 kg male NZW rabbits killed by a blow to the neck and perfused by a Watson-Marlow pump with Holman's solution¹⁷ ($p\text{O}_2$ 450–550 mmHg at 35°C). Phasic pressure variations were damped to $<5\%$ mean pressure by an air-filled compliance chamber. Zero pressure corresponded to zero flow. Experiments were conducted after 60–90 min stabilization. Microradiographs were obtained at steady-state perfusion pressure at each flow rate studied. A $4\text{-}\mu\text{m}$ microfocal X-ray source, exposures of 8–12 s and low kilovoltage techniques (10–25 kV) were used to optimize radiographic parameters. Radiographic contrast was provided by a pulse of low concentration non-ionic contrast medium (iohexol, $100 \text{ mg iodine ml}^{-1}$, diluted in Holman's buffer) which had negligible effect on perfusion pressure ($<6\%$ fall). The buffer contained 5% dextran (of relative molecular mass 80,000) to provide viscosity (2.23 mPas) equal to the contrast solution. Haemoglobin was freshly prepared by lysis of washed human red blood cells and chromatographic purification²⁴. Measurement from microradiographs was assisted by an IBAS Kontron semi-interactive image analysis system.

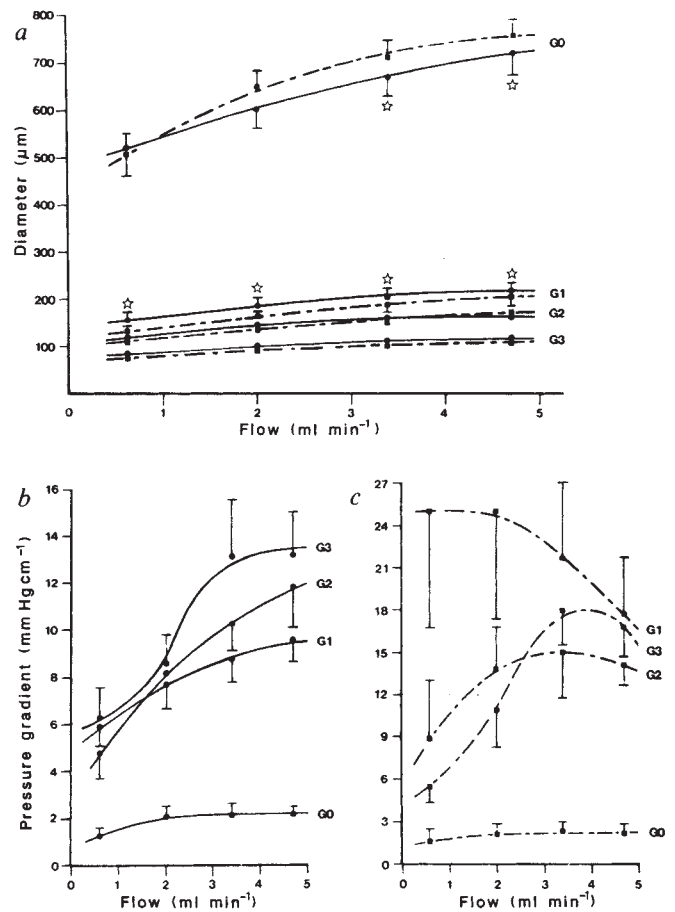


Fig. 2 *a*, Mean diameter ($\pm$ s.e.m.) plotted as a function of flow in resting beds in the absence (continuous lines) and presence (broken lines) of $1 \mu\text{M}$ haemoglobin ($n=9$ preparations). Vessel lengths were not altered by changes of flow. Significant differences induced by haemoglobin are indicated ($\star$, $P<0.05$, Student's *t*-test). *b*, *c*, Calculated pressure gradients in G0, G1, G2 and G3 in control preparations plotted as a function of flow. The curves are smoothly increasing, asymptotic functions although a point of inflexion occurs in that for G3 (panel *b*). Haemoglobin increased the pressure gradients in G1–G3 (panel *c*, note difference in scale) which then tended to decline at high flow rates, particularly in G1.

Methods. Average pressure gradients (dP/dL) were calculated from the diameter-flow data illustrated in *a*, in the absence and presence of haemoglobin. In any generation $dP/dL = 128\mu/\pi NK \sum_{j=1}^K \dot{Q}_j/D_j^4$ where K is the number of preparations analysed (that is 9), μ is the dynamic viscosity of the buffer, D_j the diameter of the *j*th vessel and N the total number of vessels in each generation. N was estimated from the microradiographs as 1, 20, 75 and 250 in G0 to G3 respectively.

recently been shown to be nitric oxide²¹, which is bound by haemoglobin²². In buffer-perfused preparations, haemoglobin inhibits EDRF activity induced by increased flow¹² and was used experimentally for this purpose; it does not increase vascular tone in the absence of endothelium^{18,23,24}.

In intact ear preparations, inhibition of basal EDRF activity significantly constricted the first generation of branch vessels (G1) at all flow rates, slightly constricted the second and third generations (G2 and G3) in some preparations, but generally dilated the central ear artery (G0) (Figs 1, 2*a*). The paradoxical dilation of G0 can be attributed to the overriding effect of a rise in perfusion pressure caused by 'downstream' constriction. These findings provide direct evidence that basal EDRF activity, by inhibiting spontaneous tone, influences the behaviour of resistance vessels *in situ*. They show that its influence is most

evident in G1 in this preparation. It has been suggested that homeostatic responses to perturbations in pressure and flow within a vascular network will occur at sites where there is the greatest 'mechanical leverage'²⁵. EDRF may function as just such a local regulator of flow: vessels in generation G1 divide without anastomosing (Fig. 1) and topographically are therefore 'feed vessels', well sited to control the distribution of flow.

Previous workers have assumed that a simple power law relates local values of diameter (D) and flow ($\dot{Q}$) in a microvascular bed, that is, $\dot{Q} = aD^M$ where a is a constant²⁶. In the present study an expression of the form $\dot{Q} = aD^M + b$ was fitted to the diameter-flow relationships in each generation by a least-squares minimization procedure²⁷; the constant b was inserted because vascular diameters are not zero at zero flow²⁸. This expression was found to describe the aggregate behaviour of

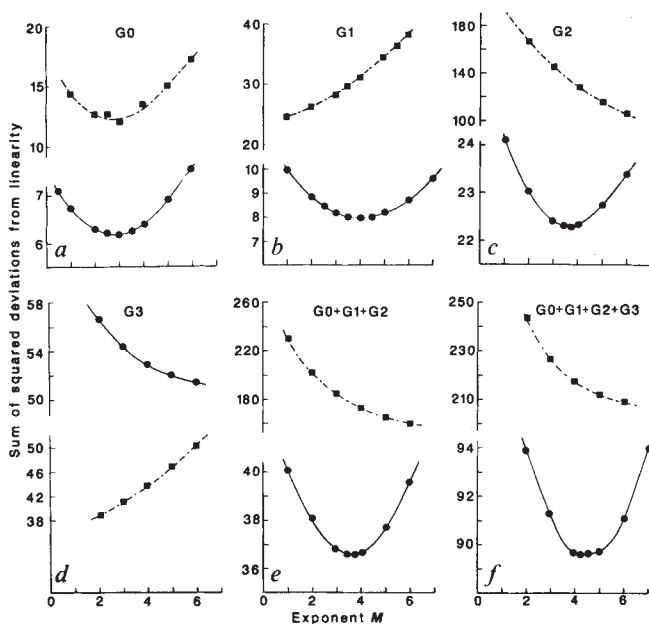


Fig. 3 Sums of squared deviations from linearity (SSD) after transformation of the diameter-flow data of Fig. 2a, according to the relationship $\dot{Q} = aD^M + b$, plotted against M in the absence and presence of haemoglobin. Using data for each generation ($n=9$), minima were found in G0 (a), G1 (b) and G2 (c) but not in G3 (d), and also in pooled data G0-2 or G0-3 (e,f). In G0 haemoglobin did not abolish the minimum but increased SSD; in G1, G2, G0-2 and G0-3 haemoglobin abolished the minima. The M values of the minima in the absence of haemoglobin were $G0=3$, $G1=4$, $G2=3.8$, $G0-2=3.75$, $G0-3=4.1$. No minima could be found in the presence of 5-HT (either 0.1 or 1 μM , Fig. 4). Confidence limits for the estimates of M were sought by alternatively analysing data for individual vessels from the experiments of Fig. 2 and from the control data of Fig. 4 (total $N=21$). Of all vessels, 55% exhibited minima. From this subpopulation estimates of M were 3.8 ± 0.6 in G0, 4.0 ± 0.3 in G1, 4.0 ± 0.4 in G2, 3.6 ± 0.4 in G3 and 3.9 ± 0.2 in G0-3 (mean $\pm$ s.e.m.).

Methods. Data were fitted by a microcomputer program to the equation $\dot{Q} = aD^M + b$. The computer chose a value of the exponent M and fitted a straight line through the four data points for each vessel by linear regression. These lines were then transformed to $Y = X$ where $Y = (\dot{Q} - b)/a$ and $X = D^M$, to normalize the data and render the analysis independent of vessel size and therefore branch generation. The sum of squared deviations (SSD) from linearity for the line $Y = X$ was then computed for each vessel, and hence the total SSD for any subgroup of arteries could be determined. If a valid linear relationship exists between $\dot{Q}$ and D^M , the total SSD for any group of arteries should exhibit a minimum when plotted against various values of M .

the network under control conditions (Fig. 3), but did not adequately fit data obtained in the presence of haemoglobin or during pharmacological constriction of the bed. The analysis assumes that the flow in any given vessel ($\dot{Q}$) is directly proportional to the input flow to the whole bed, as would be the case if its topography could be represented by a series-parallel network. This assumption is justified because anastomoses generally occurred only in G3 and higher order generations and because mean values of pressure and flow can be predicted from such a series-parallel model²⁹, even though branching patterns influence pressure-flow relationships²⁸⁻³¹. In each preparation a mean 'spatially averaged' diameter for each of generations G1-G3 was obtained by meaning data from five separate vessels, and G0 was measured at a point midway between the base and tip of the ear. Figure 3 illustrates best estimates of the exponent M in different generations singly and in combination. In G0, where passive distensive effects may contribute, it was 3.0; in

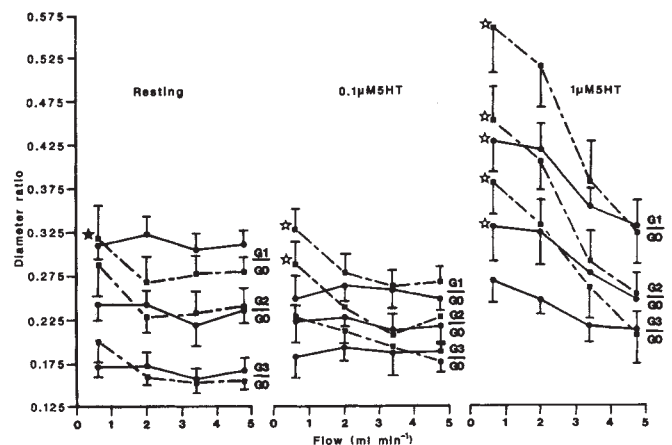


Fig. 4 The geometrical similarity of internal vessel diameters at each flow rate in generations G0 through G3 was compared by arbitrarily normalizing the diameters in G1, G2 and G3 as a fraction of G0 for each preparation. Data were obtained in unstimulated preparations (left, $n=9$), and after pharmacological constriction with 5-hydroxytryptamine (5-HT) 0.1 μM (middle, $n=5$), or 1 μM (right, $n=7$), in the absence and presence of haemoglobin. Diameter ratios at each flow rate were compared non-parametrically (to avoid the assumption of their being normally distributed) by the Friedman two-way analysis of variance by ranks technique³⁵, the null hypothesis being that they were independent of flow rate. The figure shows that this hypothesis was satisfied in resting and moderately constricted vessels but that the graphs tend to depart from the horizontal when EDRF is inhibited by haemoglobin and at higher levels of constriction ($\star$, $0.1 > P > 0.05$; $\star\star$, $P < 0.05$). If diameters were alternatively normalized with reference to G1, slopes deviated significantly ($P < 0.05$) from the horizontal in the presence of haemoglobin in G0 and G3 (but not G2) in resting vessels, and in G0, G2 and G3 in the presence of 0.1 μM 5-HT.

G1, where EDRF activity is maximal (Fig. 2), it was 4.0; in G2 it was 3.8; in G3 no minimum was found; pooled data from all generations gave a value of 4.1. These findings suggest that the influence of EDRF on flow-related vasomotion in these vessels may be described by a law of the form $\dot{Q} = aD^4 + b$ under normal control conditions. Although the experiments were conducted with non-pulsatile flow, it is unlikely that this conclusion would be different with pulsatile flow, which enhances EDRF release^{11,12}. EDRF activity under these experimental conditions of steady flow appears to be maximal, in that 50 μM glyceryl trinitrate, which may be regarded as an EDRF analogue, did not reduce perfusion pressure in control preparations.

A fourth-power law has important implications. It implies that the pressure gradient in each vessel and thus the absolute pressure drop across it will become independent of flow at high flow rates: in Poiseuille flow the pressure gradient (dP/dL) is proportional to $\dot{Q}/D^4$, so that in vessels where $\dot{Q} = aD^4 + b$, dP/dL will be proportional to $a/(1 - b/\dot{Q})$ and asymptotically approach a constant value at high flows (see Fig. 2b). This finding is supported by experimental data showing near constancy of pressure gradient across the proximal vessels (down to 70 μm) of the cat mesentery subjected *in vivo* to changes in central artery pressure^{30,31}. The general conclusion may thus be relatively independent of the complexities of pulsatile flow, dependency of blood viscosity on shear-rate and the spatial variation of haematocrit which occur *in vivo*³². A second implication of a fourth power law is that power losses become proportional to flow, since they are proportional to $dP/dL \times Q$ in Poiseuille flow and dP/dL becomes constant at high flow. The energy expended in delivering a given volume of perfusate to the terminal elements of the vascular bed under control conditions will then become independent of flow rate, instead of being directly proportional to it as in a rigid tube. As the fourth-power

law holds only in the presence of EDRF activity, EDRF appears to provide a mechanism which limits the work of perfusion.

EDRF also regulated the network in such a way that the relative change in diameter in each vessel was the same when flow was altered. Figure 4 illustrates this constancy of relative diameter, which was lost in the presence of haemoglobin and at high levels of constriction. Geometrical similarity of vessel diameters was maintained in partially constricted preparations and in G3; in neither of these was a fourth-power relationship between diameter and flow demonstrated, so it appears not to depend on this relationship. It implies geometrical similarity in respect of hydraulic resistance, which varies as L/D^4 and thus in practice on diameter alone (D) (as vessel lengths (L) remained constant). It follows that the spatial distribution of flow will be independent of flow rate. EDRF activity thus appears to interact with the myogenic and passive properties of the individual vessels, even in a partially constricted network, to coordinate the distribution of flow.

Murray³³ suggested that the optimum topography of a vascular network (under conditions of invariant flow) is one where flow is related to the cube of diameter because this minimizes power losses written as the sum of viscous power losses and a factor proportional to the volume of a section of artery ($W = aQ^2/d^4 + bd^2$). Such a third-power law has indeed been shown to describe the regional dependence of flow on diameter in arterioles and arteries (6–108 μm in diameter) of the rat cremaster muscle²⁶, and to govern slow adaptation to altered flow in canine carotid artery³⁴. Murray's analysis did not however take into account the possibility of flow-dependent dilation, nor did it address the effect of acute changes in flow—where we have shown that EDRF coordinates responses according to a fourth-power law. Whereas a third-power law minimizes power losses, a fourth-power law minimizes energy loss per unit volume which is proportional to Q/D^4 in Poiseuille flow. EDRF could also provide the essential functional signal for the regulation of arterial calibre in the long-term, however, as a third-power law normalizes shear stress ($\propto Q/D^3$) and EDRF activity is related to shear^{11,12}.

We thank Mrs E. D. Walker and Mrs N. L. Kimm for secretarial assistance, Mr M. Stanton, Mr N. Garrahan and Professor E. D. Williams, for provision of image analysis facilities, Dr D. A. Hutton for viscosity measurements and Professor A. H. Henderson and Dr A. C. Newby, for critical comment on the manuscript. The work was supported by the British Heart Foundation.

Received 24 April; accepted 16 July 1987.

- Bayliss, W. M. *J. Physiol., Lond.* **28**, 220–231 (1902).
- Johnson, P. C. in *Handbook of Physiology Vol II Vascular Smooth Muscle* (eds Bohr, D. F., Somlyo, A. P. & Sparks, H. S.) 409–442 (Am. Physiol. Soc., Bethesda, Maryland, 1980).
- Folkow, B. *Circ. Res.* **15**, Suppl. 1, 279–287 (1964).
- Guyton, A. C. cited in Koch A. R. *Circ. Res.* **15**, Suppl. 1, 269–279 (1964).
- Schretzenmayr, A. *Pflüger's Arch. ges. Physiol.* **232**, 743–748 (1933).
- Hilton, S. M. *J. Physiol., Lond.* **149**, 93–111 (1959).
- Gerova, M. *et al. Basic Res. Cardiol.* **76**, 503–507 (1981).
- Smiesko, V., Kozik, J. & Dolezel, S. *Blood Vessels* **22**, 247–251 (1985).
- Holtz, J., Giesler, M. & Bassenge, E. *Z. Kardiol.* **72**, Suppl. 3, 98–106 (1983).
- Holtz, J., Forstermann, U., Pohl, U., Giesler, M. & Bassenge, E. *J. Cardiovasc. Pharmacol.* **6**, 1161–1169 (1984).
- Rubanyi, G. M., Romero, J. C. & Vanhoutte, P. M. *Am. J. Physiol.* **250**, H1145–H1149 (1986).
- Pohl, U., Busse, R., Kuon, E. & Bassenge, E. *J. appl. Cardiol.* **1**, 215–235 (1986).
- Kaiser, L., Hull, S. S. & Sparks, H. V. *Am. J. Physiol.* **250**, H974–H981 (1986).
- Khayutin, V. M. *et al. Acta physiol. Hungarica* **68**, 241–251 (1986).
- Furchgott, R. F. *Circ. Res.* **53**, 557–573 (1983).
- Griffith, T. M., Edwards, D. H., Lewis, M. J., Newby, A. C. & Henderson, A. H. *Nature* **308**, 645–647 (1984).
- Griffith, T. M., Henderson, A. H., Hughes Edwards, D. & Lewis, M. J. *J. Physiol., Lond.* **351**, 13–24 (1984).
- Martin, W., Furchgott, R., Villani, G. M. & Jothianandan, D. *J. Pharm. exp. Ther.* **237**, 529–538 (1986).
- Collins, P. Chappell, S. P., Griffith, T. M., Lewis, M. J. & Henderson, A. H. *J. Cardiovasc. Pharmacol.* **8**, 1158–1162 (1986).
- Davies, R. L., Flores, N. A. & Evans, K. T. *Br. J. Radiol.* **59**, 273–276 (1986).
- Palmer, R. M. J., Ferrige, A. G. & Moncada, S. *Nature* **327**, 524–526 (1987).
- Gibson, Q. H. & Roughton, F. J. W. *J. Physiol., Lond.* **136**, 507–526 (1957).
- Martin, W., Villani, G. M., Jothianandan, D. & Furchgott, R. F. *J. Pharm. exp. Ther.* **232**, 708–716 (1985).
- Edwards, D. H., Griffith, T. M., Ryley, H. C. & Henderson, A. H. *Cardiovasc. Res.* **20**, 549–556 (1986).

- Zweifach, B. W. *Circ. Res.* **41**, 380–390 (1977).
- Mayrovitz, H. N. & Roy, J. *Am. J. Physiol.* **245**, H1031–H1038 (1983).
- Colquhoun, D. in *Lectures on Biostatistics* 259–272 (Clarendon Press, Oxford, 1971).
- Gore, R. W. *Am. J. Physiol.* **222**, 82–91 (1972).
- Popel, A. S. *J. appl. Mech.* **47**, 247–253 (1980).
- Gore, R. W. *Circ. Res.* **34**, 581–591 (1974).
- Zweifach, B. W. *Circ. Res.* **34**, 843–857 (1974).
- Schmid-Shönbein, H., Fischer, T., Driessen, G. & Rieger, H. in *Quantitative Cardiovascular Studies* (eds Hwang, N. H. C., Gross, D. R. & Patel, D. J.) 353–417 (University Park Press, Baltimore, 1979).
- Murray, C. D. *Proc. natn. Acad. Sci. U.S.A.* **12**, 207–214 (1926).
- Kamiya, A. & Togawa, T. *Am. J. Physiol.* **239**, H14–H21 (1980).
- Colquhoun, D. in *Lectures on Biostatistics* 200–204 (Clarendon Press, Oxford, 1971).

Characterization of murine interleukin B by a monoclonal antibody

Paolo del Guercio, Marie-France del Guercio & David H. Katz

Division of Immunology, Medical Biology Institute, 11077 North Torrey Pines Road, La Jolla, California 92037, USA

Interleukin B (IL-B) (ref. 1), formerly termed BEF (B-cell-derived enhancing factor)^{2,3} or IL-B4 (ref. 4), was originally described as a non-immunoglobulin regulatory factor spontaneously produced by B lymphocytes and B-cell lines that enhances the *in vitro* antigen-driven antibody response of unfractionated spleen cells stimulated by thymus-dependent antigens^{2,3}. Since then we have examined the function of interleukin B in a number of immune reactions, both *in vitro* and *in vivo*⁵, and found that it inhibits the activation of suppressor T lymphocytes^{3,6–9}. We report here the production of two monoclonal antibodies (mAb) that specifically inhibit interleukin B activity. The use of these mAb in the purification and characterization of IL-B is described. IL-B from both normal and transformed B cells consists of two subunits of similar size and amino-acid composition. The structure of interleukin B and its specific behaviour in biological assay distinguish it from many other known lymphokines.

Murine IL-B was obtained from the non-immunoglobulin-expressing B-cell line P3-X63-Ag.8.653¹⁰ and rat IL-B from the non-immunoglobulin-secreting B-cell line IR983F¹¹. Partially purified IL-B of mouse and rat origin was used to immunize rats and mice respectively. Spleen cells from these animals were fused with the B-cell line P3-X63-Ag8.653. Hybridoma supernatant fluids were screened for anti-IL-B activity by solid-phase assay (ELISA) and supernatant from positive hybridoma clones was assayed for its capacity to inhibit IL-B enhancement of antibody synthesis by spleen cells¹². Subcloning of selected anti-IL-B hybridomas produced two lines: MF-414.XIII (rat IgM) and MC-H23 (mouse IgG₁). Antibodies from both lines react in solid-phase assay (ELISA) with mouse and rat IL-B, and they were used interchangeably in the experiments dealing with murine IL-B. As illustrated in Fig. 1a, as little as 100 ng of purified MC-H23 antibody completely reversed the inhibitory activity exerted by IL-B on the activation of the suppressor T-cell hybrid T2D4-P. Comparable concentrations of an unrelated mouse mAb of the same isotype failed to inhibit the IL-B activity. Moreover, the same dose of MC-H23 inhibited IL-B enhancement of the antibody response (Fig. 1b). A similar inhibition of IL-B regulatory activities was found using purified MF-414.XIII antibody.

Both MC-H23 and MF-414.XIII mAb were used successfully partially to purify IL-B (see legend to Fig. 2). Purified biologically active IL-B was recognized by MC-H23 mAb in ELISA (Fig. 2, panel A). SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis (14.5% polyacrylamide) showed that IL-B ran as two bands of relative molecular mass (M_r) 16,500 and 15,000 (16.5K, 15K). To test whether both proteins could be detected

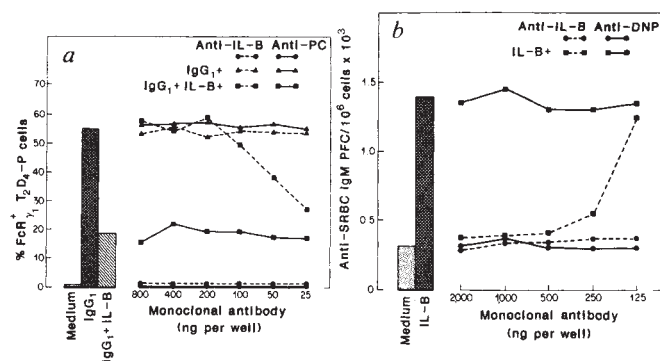


Fig. 1 Anti-IL-B mAb *a*, reverses the inhibitory activity exerted by IL-B on the activation of the suppressor T-cell hybrid T2D4-P, *b*, inhibits IL-B enhancement of antibody synthesis.

Methods. *a*, Lou rats and BALB/c mice were injected intraperitoneally with 10 μg (emulsified in 0.1 ml of complete Freund's adjuvant) partially purified mouse or rat (respectively) IL-B¹², and were boosted four times, at 4-week intervals, with 10 μg of IL-B adsorbed on 4 mg alum. Five days after the last injections, spleen cells were fused with the P3-X63-Ag8.653 line. Hybridoma supernatants (96-well plates) were screened for their capacity to react in a solid-phase assay with IL-B purified by affinity chromatography on an anti-IL-B rabbit IgG immunoadsorbent. Positive hybridomas were transferred into 24-well plates and supernatants (100 μl) from these were tested for their capacity to inhibit IL-B enhancement of *in vitro* antigen-driven IgM synthesis by spleen cells (10^6 cells per 0.1 ml)¹². Two hybridomas that reproducibly inhibited IL-B activity were subcloned, resulting in two monoclonal antibodies: MF-414.XIII, which secretes rat IgM antibodies, and MC-H23, which secretes mouse IgG1. T2D4-P cells were generated by subcloning the suppressor T-cell hybrid T2D4, which can be activated by mouse IgG₁ to express $\text{Fc}\gamma\text{R}_1$ (activation phase) and subsequently releases an IgG₁-specific suppressor factor (effector phase). It has been shown that IL-B is highly effective in inhibiting the activation of T2D4-P cells, but it is unable to inhibit the release of suppressor factor by already activated cells^{5,7}. T2D4-P cells (5×10^5 cells in 0.1 ml) were cultured either in medium alone or in the presence of 20 μg murine IgG₁ per well (anti-DNP mAb 109.3) alone or together with 1.2 μg affinity-purified IL-B per well. Various concentrations of anti-IL-B mAb MC-H23 or anti-phosphorylcholine (PC) mAb 24B5, both of the mouse IgG₁ isotypes, were added to the culture. After a 2-h incubation the cells were assayed for $\text{Fc}\gamma\text{R}_1^+$ cells by rosette formation with dinitrophenylated sheep red blood cells (DNP-SRBC) as described⁹. Similar results were obtained with the anti-IL-B mAb MF-414.XIII. *b*, Spleen cells from DBA/2 mice (10^6 in 0.2 ml) were cultured in the presence or absence of 2.4 μg affinity-purified IL-B per well and various concentrations of anti-IL-B mAb MC-H23 or anti-DNP mAb 109.3, both of the mouse IgG₁ isotype. Cultures were stimulated with SRBC, and direct anti-SRBC plaque-forming cells (PFC) were determined on day 5 of culture⁶. Similar results were obtained with the anti-IL-B mAb MF-414.XIII.

by monoclonal antibody, purified IL-B was electrophoresed and transferred to nitrocellulose paper. Figure 2 shows that MC-H23 mAb reacts with both 16.5K and 15K proteins but not with the proteins used as molecular weight markers (panel *c* and panel *b*). The amino-acid composition of the individual protein bands was determined after electroelution from the gel and was found to be closely similar. An amino-acid sequence could not be determined because the N-terminus of the proteins was blocked.

We then compared the size of IL-B produced by normal and unstimulated B lymphocytes with IL-B produced by a transformed cell line, B lymphocytes being purified as previously described² and cultured at a concentration of 10^7 cells per ml in medium containing 5% heated fetal calf serum (FCS). IL-B was purified by affinity chromatography and RP HPLC (see Fig. 2 legend) and chromatography fractions were screened in ELISA using the anti-ILB mAb MC-H23. Only one sample

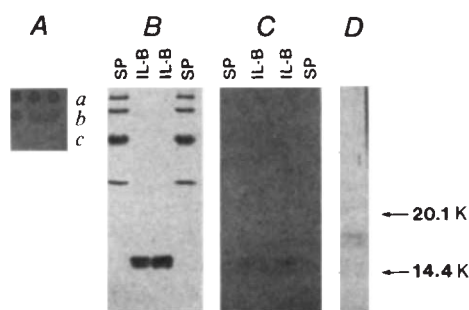


Fig. 2 Biochemical analysis of IL-B purified from a B-cell line or from normal B lymphocytes. Panel A, ELISA analysis of IL-B purified from a B-cell line. *a*, 15 ng or *b*, 7.5 ng RP-HPLC-purified IL-B or *c*, 15 ng mouse Ig were absorbed onto nitrocellulose paper (Millipore). The paper was blocked for 1 h at room temperature with 1% bovine serum albumin, washed with 50 mM Tris, pH 7.9, 150 mM NaCl (TBS) and then incubated for 2 h at room temperature with horseradish peroxidase-conjugated MC-H23 mAb in TBS containing 0.05% Tween 20. After washing, the paper was developed with a peroxidase substrate solution. This was prepared by mixing 2 ml 4-chloro-1-naphthol (3 mg ml⁻¹) in methanol with 10 ml of TBS plus 0.01 M imidazole and then adding 30% hydrogen peroxide (5 μl). Panel B, SDS-PAGE (14.5% polyacrylamide) analysis of IL-B purified from a B-cell line. Electrophoresis of purified IL-B and molecular weight standard proteins (SP) was done under reducing conditions (2-mercaptoethanol), and proteins were detected by silver staining. Panel C, Western blotting of IL-B purified from a B-cell line. Electrophoresis as described in panel B; the protein bands were transferred to nitrocellulose paper using a trans-blot cell (Bio-Rad; 18 h at 10 mV) and visualized with horseradish peroxidase-conjugate MC-H23 mAb. Panel D, SDS-PAGE (14.5% polyacrylamide) of IL-B purified from normal B lymphocytes.

Methods. Serum-free culture supernatant fluids were applied to anti-IL-B mAb-coupled Sepharose columns and IL-B activity eluted with 0.2 M glycine, pH 2.4, buffer. Purified IL-B was subjected to reverse-phase high pressure liquid chromatography (RP-HPLC) on an aquapore octyl column (250 $\times$ 4.6 mm) (Brownlee, California) and eluted with a linear gradient of 40 to 60% acetonitrile over 30 min at 2 ml per min. IL-B activity was eluted at 47–50% acetonitrile.

(eluted at approximately 50% acetonitrile) was positive. All samples were tested for their capacity to inhibit the activation of the suppressor T-cell hybrid T2D4-P, but only the sample positive in the ELISA showed inhibitory activity (55% inhibition). SDS-PAGE analysis (Fig. 2, panel D) showed a comparable M_r for active IL-B purified from normal and unstimulated B lymphocytes and IL-B purified from the continuous B-cell line. A biologically active IL-B purified from the B-cell line of rat origin IR983F¹¹ also interacts with the mAb MC-H23 and has a molecular size of 16K–18K (data not shown).

The specific activity of IL-B purified from the murine B-cell line was approximately 4,000 times higher than that of the crude supernatant fluid and 50 times higher than that of affinity-purified (MC-H23) IL-B. Figure 3*a* shows that 50 ng of IL-B purified by RP-HPLC reversed the activation of the suppressor T-cell hybrid T2D4-P. Figure 3 shows that 12.5 ng of IL-B resulted in a fivefold enhancement of the antigen-driven *in vitro* IgM antibody response.

Purified IL-B did not show IL-1, IL-2, BSF₁ (IL-4), BCGFII (IL-5) or interferon- γ (IFN γ) activity in specified biological assays. In particular, IL-B was tested for its capacity to increase the response of C3H/HeJ thymocytes to phytohaemagglutinin (IL-1), to sustain the growth of the T-cell line CTLL-2 (IL-2), to induce proliferation of NK cells (IL-2, BSF₁) or BCL₁ cells (BDGFII), to induce Ia expression on B cells (BSF₁) or on WEHI-3 cells (IFN γ), and to activate macrophages for non-specific tumoricidal activity (IFN γ) (P. del Guercio *et al.*, in preparation). Neither of the monoclonal antibodies described

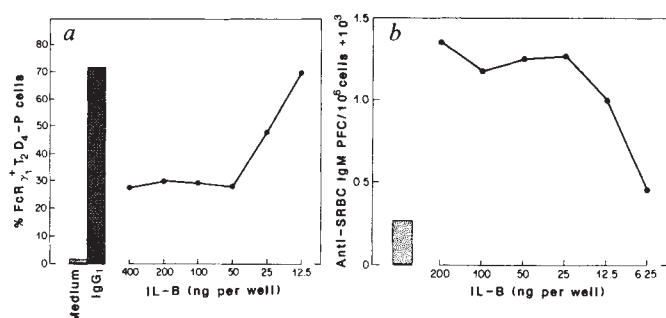


Fig. 3 RP-HPLC-purified IL-B *a*, inhibits activation of the suppressor T-cell hybrid T2D4-P; and *b*, enhances *in vitro* antibody response.

Methods. *a*, VT2D4-P cells (5×10^5 cells in 0.1 ml) were cultured either in medium alone or in the presence of 20 μg IgG₁ (anti-DNP mAb 109.3) per well and various concentrations of RP-HPLC-purified IL-B. After 2 h incubation the cells were assayed for $\text{Fc}\gamma\text{R}^+$ cells by rosette formation. *b*, Spleen cells from DBA/2 mice (10^6 cells in 0.2 ml) were cultured in the presence of various concentrations of RP-HPLC-purified IL-B. Cultures were stimulated with SRBC, and numbers of direct anti-SRBC PFC were determined on day 5.

in this report reacts in ELISA with IL-1, IL-2 or interferon γ or with monoclonal mouse antibodies of various isotypes (results not shown). Moreover, on the basis of the amino-acid composition (Table 1), IL-B can be distinguished from other already sequenced lymphokines in that IL-B has no cysteine, a high content of alanine and lysine and a low content of leucine, aspartic acid and asparagine. These findings confirm earlier results⁵ suggesting that IL-B is a unique lymphokine.

This characterization of a novel lymphokine is important for three reasons. First, these studies clearly establish the existence of non-immunoglobulin regulatory molecules produced spontaneously by B cells. Second, this is the first molecule to reduce the activation of suppressor T cells. In this regard, partially purified IL-B has been used successfully in mice to reduce the activation of suppressor T cells during acute graft-versus-host reactions (unpublished results). Third, IL-B could be important in the regulatory network that ensures the physiological dynamic equilibrium of the immune system. By inhibiting the activation of developing clones of suppressor T cells and therefore permitting the expansion of B-cell clones that are reacting against environmental epitopes, IL-B could promote the production of constant levels of immunoglobulin molecules in the absence of

external stimuli¹. The availability of an anti-IL-B monoclonal autoantibody, which is a powerful inhibitor of IL-B activity, will allow us and other interested groups to examine the role of IL-B *in vivo*.

This work was supported by INSERM, by NIH and by a grant-in-aid from QUDEL, La Jolla. We thank Drs John F. Marcelletti and Donald E. Mosier for discussion, Drs Tony E. Hugli and Mike Ballard for amino-acid composition analysis and Janet Czarnecki for preparing the manuscript.

Received 16 June; accepted 30 July 1987.

- del Guercio, P., Marcelletti, J. F. & Katz, D. H. *Immun. Today* **7**, 97-98 (1986).
- del Guercio, P., Poirier, M.-F. & Brugere, S. *Cell Immun.* **72**, 316-325 (1982).
- del Guercio, P., Brugere, S. & Poirier, M.-F. *Cell Immun.* **73**, 360-364 (1982).
- del Guercio, P., Marcelletti, J. F., Zuberi, R. I., del Guercio, M.-F. & Katz, D. H. *J. cell Biochem. Suppl.* **10A**, 84 (1986).
- del Guercio, P. & Katz, D. H. *CRC Crit. Rev. Immun.* **6**, 185-212 (1986).
- del Guercio, P. & Poirier, M.-F. *J. Immun.* **131**, 1283-1288 (1983).
- del Guercio, P., Fridman, W. H., del Guercio, M.-F. & Katz, D. H. in *Advances in Gene Technology: Molecular Biology of the Immune System* (eds Streilen, J. W. et al.) 175-176 (Cambridge University Press, 1985).
- del Guercio, P., del Guercio, M.-F., Fridman, W. H. & Katz, D. H. *J. Immun.* **138**, 3295-3299 (1987).
- del Guercio, P., Zanetti, M., del Guercio, M.-F. & Katz, D. H. *J. Immun.* **134**, 3926-3933 (1985).
- Kearney, J. F., Radbruch, A., Liesegang, B. & Rajewsky, K. *J. Immun.* **123**, 1548-1550 (1979).
- Bazin, H. in *Protides of the Biological Fluids* (ed. Peeters, H.) 615-618 (Pergamon, New York, 1982).
- del Guercio, P., del Guercio, M.-F., Zanetti, M. & Katz, D. H. *Cell Immun.* **98**, 337-340 (1986).

HLA-restricted recognition of viral antigens in HLA transgenic mice

Femia Kievits*, Pavol Ivanyi*, Paul Krimpenfort†, Anton Berns† & Hidde L. Ploegh‡

* Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, Department of Immunohematology and Laboratory of Experimental and Clinical Immunology, University of Amsterdam, Plesmanlaan 125, 1066 CX Amsterdam

† Department of Genetics, The Netherlands Cancer Institute and Department of Biochemistry, University of Amsterdam and

‡ Department of Cellular Biochemistry, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Cytotoxic T lymphocytes (CTL) recognize antigen in the context of the class-I products of the major histocompatibility complex (MHC)¹. The extensive polymorphism of class-I molecules is thought to be linked to their capacity to present a large variety of foreign antigens¹. Whether a single T-cell receptor (TCR) recognizes two separate epitopes (the foreign antigen and an epitope on MHC molecules^{2,3}), or a single epitope resulting from the combination of a foreign antigen and an MHC molecule, has not yet been resolved⁴⁻⁶. In view of the differences between species in primary structure of histocompatibility antigens, it might be predicted that the TCR repertoire would evolve in concert with the diversity of MHC antigens. The mouse and human TCR repertoire would be optimally adapted to engage in productive interactions only with mouse (H-2) and human (HLA) MHC antigens respectively, especially if the more conserved features of histocompatibility antigens, in addition to foreign antigen, were seen by the TCR⁷. Alternatively, only the most variable segments of MHC antigens might be engaged in antigen presentation and thus in interaction with the TCR⁸. In that case, interaction between MHC plus antigen and the TCR might not necessarily be limited by species-specific features. By analysis of the T-cell response against virus-infected cells in HLA-B27/human β_2 -microglobulin double transgenic mice, we report here that the mouse T-cell repertoire is perfectly capable of using the human HLA-B27 antigen as a restriction element.

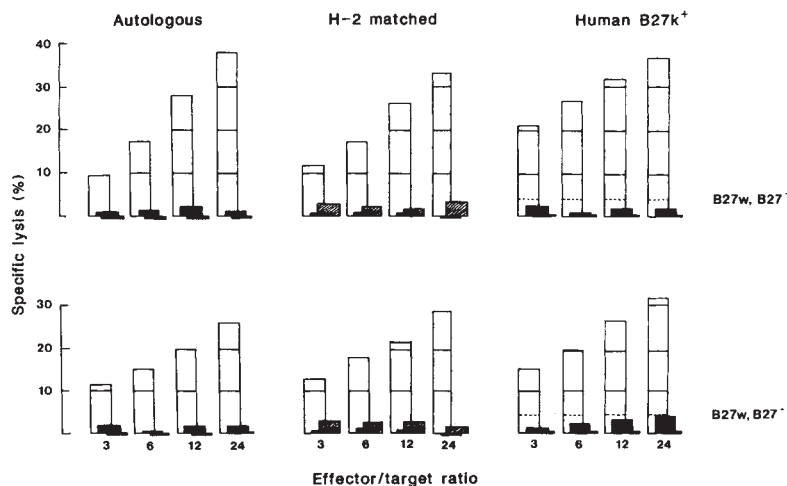
The HLA-B27 specificity can be subdivided in a number of subtypes based on the reactivity of allospecific and HLA-

Table 1 Amino-acid composition of mouse IL-B*

Amino acid	%
Aspartic acid/asparagine	6.2
Threonine	5.5
Serine	9.7
Glutamic acid/glutamine	9.0
Glycine	8.3
Alanine	9.0
Valine	6.9
Methionine	1.4
Isoleucine	4.1
Leucine	6.2
Tyrosine	4.1
Phenylalanine	6.2
Histidine	2.8
Lysine	11.7
Arginine	5.5
Proline	3.4
Cysteine	0

* Results are given for the 16.5 K protein. The 15 K protein has a similar composition.

Fig. 1 Cytotoxic T lymphocytes (CTL) specific for Sendai virus (upper panels) and influenza virus (lower panels) were generated in double transgenic mice¹⁶ and tested on autologous targets (left), H-2 matched non-transgenic targets (centre), and human HLA-B27K positive cells (right). In each case, CTL were tested on target cells infected with the virus against which the mouse was immunized (open bars), as well as on cells infected with the irrelevant virus (solid bars; influenza virus in the case of Sendai-specific CTL; Sendai virus in the case of influenza-specific CTL), and uninfected target cells (hatched bars), at the effector/target ratios indicated. Only those virus-infected targets expressing HLA-B27K or matched for autologous H-2 antigens are lysed. Lysis of human HLA-B27W⁺ or HLA-B27⁻ cells was below 4% specific release (indicated by the dashed line; rightmost panels) for virus-infected as well as uninfected cells (data not shown).



Methods. Double transgenic mice were primed *in vivo* by intraperitoneal injection of 1,000 haemagglutination units (HAU) of Sendai virus or 1,500 HAU of influenza virus A/X79 (FLU) in 0.5 ml of phosphate-buffered saline (PBS). Seven days after Sendai virus priming or 21 days after influenza virus priming, mice were killed and their spleen cells (5×10^7) stimulated with 10^7 virus-infected, irradiated (2,000 rad) autologous spleen cells in 45 ml of Iscove's modified Dulbecco's medium, supplemented with 10% (v/v) fetal calf serum (FCS), 100 IU ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and 5×10^{-5} M 2-mercaptoethanol. Autologous lymph node cells were used as target cells. Stimulator cells were incubated with virus (300 HAU per 10^7 cells) in 1 ml of medium for 1 hour at 37 °C. Cells were washed twice and influenza-virus-infected stimulator cells were further incubated for three hours before irradiation. After five days, the responding CTL were collected and tested in the ⁵¹Cr-release assay on Sendai-virus-coated (1 h) or influenza-virus-infected (18 h) target cells. Murine and human lymphocytes stimulated for three days with 2.5 μ g ml⁻¹ concanavaline A were used as target cells. Infection of target cells was performed as described for stimulator cells. Responding CTL were incubated with ⁵¹Cr-labelled target cells at the effector/target ratio indicated. The percentage of specific release was calculated as (release in the presence of CTL - spontaneous release)/(release in 10% Triton X-100 - spontaneous release) $\times$ 100%. The spontaneous release in medium from labelled target cells ranged from 10% to 20% of the maximal release produced by Triton X-100. Human peripheral blood HLA-typed lymphocytes were isolated by flotation on Ficoll-Isopaque and stored in liquid nitrogen until used. Human cells were cultured in the presence of 10% normal human pooled serum. HLA types of the human cells were: HLA-A2, A28; B27K; Cw2; (B27K), HLA-A2, A28; B27W; Cw2; (B27W), and HLA-A2, A28, B5, B12; Cw4; (B27-). Bars, mean percentage of specific lysis of target cells at the given effector/target ratio measured in triplicate.

restricted, virus-specific CTL^{9,10}. The genes encoding two distinct HLA-B27 subtypes (HLA-B27W and HLA-B27K), which together make up the vast majority of HLA-B27-positive individuals of caucasoid origin¹⁰, have been cloned¹¹ and expressed in mouse and human cells by transfection¹². The products encoded by these two genes differ by only three amino acids^{11,13}. In the course of transfection experiments into murine recipient cells it became apparent that both the human β_2 -microglobulin (β_2m) and the HLA-B27 gene are required to obtain a high surface expression of HLA-B27 antigens¹². Only in such double transfectants, authentic HLA-B27 molecules, consisting of both human subunits, rather than hybrid products consisting of HLA-B27 and murine or bovine¹² β_2m are expressed at the cell surface. This situation is in contrast to HLA-A2 and HLA-B7 heavy chains, which readily associate with murine β_2m in transfection experiments^{12,14,15}. We therefore constructed two sets of transgenic mice, one expressing the HLA-B27K gene, and the other the human β_2m gene. Crosses between mice of these two types produce animals expressing both human subunits at the cell surface and thus authentic HLA-B27 antigens, as verified by cytofluorimetry and biochemical analysis¹⁶. The amount of HLA-B27 antigen present in spleen cells from double transgenic mice is comparable to that found in Epstein-Barr virus (EBV)-transformed human lymphoid cell lines (unpublished observations). To confirm and extend these findings it was shown that human anti-HLA-B27K specific CTL⁹ can efficiently lyse double transgenic murine target cells. Lysis of B27K positive double transgenic murine target cells was observed (29% specific release at effector/target ratio of 7:1), whereas control murine cells of HLA-B27-negative human cells were not lysed (data not shown).

HLA-B27 molecules have been characterized in humans as effective restriction elements for anti-viral CTL^{9,10,17}. To examine whether HLA-B27 as expressed in double transgenic mice would function as a restriction element, mice were primed *in vivo* with either influenza virus A/X79 or Sendai virus, and the secondary

virus-specific CTL response *in vitro* was measured by lysis of virus infected autologous, H-2 matched, or human cells. Figure 1 shows that antiviral CTL, obtained *in vitro* from double transgenic mice after priming *in vivo*, were able to kill virus-infected HLA-B27K-positive human target cells, whereas no lysis was observed of non-infected cells, or virus-infected cells expressing HLA-B27W, HLA-B5 or HLA-B12 antigens. A second set of mice were used for a similar experiment, shown in Fig. 2. The CTL populations obtained also lysed virus-infected H-2-matched target cells, derived from non-transgenic (CBA $\times$ C57B1/6) F₁ animals. These mice possess the H-2^b and H-2^d haplotypes present in the strain combination in which the transgenic mice were generated¹⁶. Thus, in double transgenic mice both murine and human class-I antigens are used as restriction elements. The specificity of the CTL for the virus with which the mouse was immunized is apparent from their lack of reactivity with target cells infected with the irrelevant virus: anti-influenza CTL do not lyse Sendai-virus-infected target cells nor do anti-Sendai-virus CTL kill influenza infected targets.

Recognition of Sendai- and influenza-virus-infected cells was restricted by the HLA-B27 subtype in the context of which immunization had taken place. Hence, murine T-cells can distinguish the same HLA-B27 subtypes—differing by only three amino acids—as can human alloreactive and anti-viral CTL^{9,10}. Note that this distinction is almost absolute. Thus, in both virus systems double transgenic mice were perfectly able to generate anti-viral HLA-B27-restricted CTL that specifically killed virus-infected HLA-B27-positive human cells.

A number of previous reports have dealt with recognition of HLA antigens by murine CTL¹⁸⁻²⁰. Because these CTL were invariably raised in animals that had not acquired tolerance to xenoeneic MHC antigens, H-2 restricted recognition of such 'foreign' antigens would be expected, and has indeed been observed¹⁸⁻²⁰. Results from such studies reflect the ability of H-2-restricted T cells to distinguish several distinct epitopes on a non-self molecule. This is of course strictly comparable to the

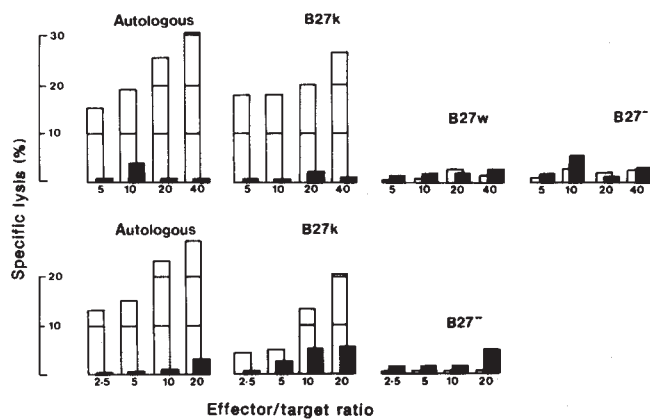


Fig. 2 Recognition of virally infected cells is restricted by the HLA-B27 subtype in the context of which was immunized. Immunization of mice, preparation of targets, and CTL assays were carried out as described in the legend to Fig. 1. In this experiment CTL were tested on infected (open bars) and uninfected (solid bars) targets only, having established that such CTL are virus-specific (Fig. 1). Bars indicate percentage of specific lysis of target cells by the given effector to target ratio measured in triplicate. The top panels depict the response of anti-influenza CTL on autologous, HLA-B27K-matched or mismatched targets, as indicated. The bottom panels show the results for the anti-Sendai response on autologous, HLA-B27K-matched, and HLA-B27 negative targets. The HLA-types of the targets were as described in the legend to Fig. 1.

ability of T cells to discriminate between epitopes on other 'foreign' antigens such as viral glycoproteins. In the experiments reported here, introduction into the germ line of all the information necessary to generate human HLA-B27 antigens has resulted in tolerance and has allowed us to test whether HLA-B27, next to the mouse's own class I MHC molecules, might serve as a restriction element for murine cytotoxic T cells. Our results show that this is indeed the case. It should be emphasized that authentic human HLA antigens are generated only when the genes encoding both human subunits are present, in contrast with the hybrid antigens produced if only HLA heavy-chain genes are used to construct transgenic animals²¹. In the latter case, a direct comparison of cells from transgenic mice with human cells as either targets or effectors should take into account the fact that human and mouse β_2m show only 70% similarity, and that hybrid molecules are distinct, at least serologically, from the parental molecules^{22,23}.

Our results strongly suggest that selection of the T-cell receptor repertoire is not appreciably affected by conserved regions of the MHC class-I molecule that contain 'species-specific' residues (for example the α -3 domain, or β_2m). The more conserved areas may merely serve to position the antigen-presenting segments of class-I antigens in the correct configuration in relation to the T-cell receptor. It is more likely that indeed the most variable regions of the class-I molecule are directly involved. Although some species-specific residues may be identified in these regions, they obviously still allow the selection of such T-cell clones that can use HLA-B27 as a restriction element.

In humans, several diseases show a strong association with the HLA-B27 antigen. At present, it cannot be decided whether this association is due to holes in the repertoire, resulting from the need for tolerance to HLA-B27 in HLA-B27 positive individuals, or alternatively, by unique properties of HLA-B27 as an antigen presenting element. The experiments reported here are a first step towards the validation of HLA-B27 transgenic mice as a possible tool in the study of HLA-disease associations.

We thank Anneloes Beenders and Esther Koorneef for technical assistance and Gea Meijerink for typing the manuscript. This work was supported by The Netherlands Organization for the Advancement of Pure Research (ZWO) through the Foundation for Medical Research (MEDIGON), and the 'Commissie voor Reumaonderzoek'.

Received 4 June; accepted 13 August 1987.

1. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51-77 (1979).
2. Epstein, R. *et al. J. exp. Med.* **163**, 759-773 (1986).
3. Cohn, M. *Cell* **33**, 657-668 (1983).
4. Matzinger, P. *Nature* **292**, 497-501 (1981).
5. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359-361 (1985).
6. Ashwell, J. D. & Schwartz, R. H. *Nature* **320**, 176-179 (1986).
7. Parham, P. R. *Immun. Today* **4**, 89-92 (1984).
8. De Waal, L., Kast, W. M., Melvold, R. W. & Melief, C. J. M. *J. Immun.* **130**, 1090-1096 (1983).
9. Breuning, M. H. *et al. Human Immun.* **5**, 259-268 (1982).
10. Toubert, A. *et al. Immunogenetics* **20**, 513-525 (1984).
11. Seemann, G. H. A., Rein, R. S., Brown, C. S. & Ploegh, H. L. *EMBO J.* **5**, 547-552 (1986).
12. Rein, R. S. *et al. J. Immun.* **138**, 1178-1183 (1987).
13. Vega, M. A., Bragado, R., Ivanyi, P., Pelaez, J. L. & Lopes de Castro, J. A. *J. Immun.* **137**, 3557-3565 (1986).
14. Herman, A., Parham, P., Weissman, S. M. & Engelhard, V. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5056-5060 (1983).
15. Bernabeu, C. *et al. J. Immun.* **131**, 2032-2037 (1984).
16. Krimpenfort, P. *et al. EMBO J.* **6**, 1673-1676 (1987).
17. Gomard, E., Sitborn, M., Toubert, A., Begue, B. & Levy, J. P. *Immunogenetics* **20**, 197-204 (1984).
18. Holtermann, M. J. & Engelhard, V. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9699-9703 (1986).
19. Achour, A. *et al. Eur. J. Immun.* **16**, 597-604 (1986).
20. Maryanski, J. L., Accolla, R. S. & Jordan, B. J. *Immun.* **136**, 4340-4397 (1986).
21. Dill, O., Koch, S. & Hammerling, G. J. *Immunobiology* **173**, 264 (1986).
22. Kievits, F. & Ivanyi, P. *Hum. Immun.* (in the press).
23. Ferrier, P., Layet, C., Caillol, D. H., Jordan, B. R. & Lemonnier, F. A. *J. Immun.* **135**, 1281-1287 (1985).

Cytotoxic T lymphocytes against a soluble protein

Uwe D. Staerz, Hajime Karasuyama* & Abigail M. Garner

Basel Institute for Immunology, Grenzacherstrasse 487, CH-4058 Basel, Switzerland

Thymus-derived (T) lymphocytes recognize antigen in conjunction with surface glycoproteins encoded by major histocompatibility complex (MHC) genes¹⁻³. Whereas fragments of soluble antigens are presented to T helper lymphocytes (T_H), which carry the CD4 antigen, in association with class II MHC molecules⁴, CD8-bearing cytotoxic T lymphocytes (CTL) usually see cellular antigens (for instance virally-encoded proteins) in conjunction with MHC class I molecules^{5,6}. The different modes of antigen presentation may result from separate intracellular transport: vesicles containing class II molecules are thought to fuse with those carrying endocytosed soluble proteins. Class I molecules, in contrast, can only pick up degradation products of intracellular proteins (see refs 7 and 8). This makes biological sense; during an attack of a virus, class I-restricted CTL destroy infected cells and class II-restricted T_H guide the humoral response to neutralize virus particles and toxins. But here we provide evidence that CTL specific for ovalbumin fragments can be induced with soluble protein, and that intracellular protein degradation provides epitopes recognized by these CTL. These findings suggest the existence of an antigen presenting cell that takes up soluble material and induces CTL.

It is well documented that fragments of cytoplasmic and membrane-associated proteins, such as influenza nucleoprotein and MHC molecules, can be seen in the context of class I molecules^{5,9-11}. But the possibility that antigens taken up from outside cells can be presented to class I-restricted T lymphocytes

* Present address: Department of Immunology, Faculty of Medicine, University of Tokyo, Tokyo, Japan.

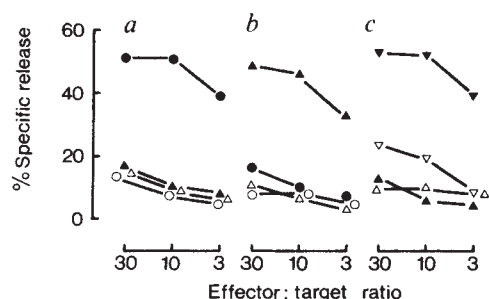


Fig. 1 Specificity of secondary *in vitro* cultures after *in vivo* immunization with soluble ovalbumin. Cultures derived from C57BL/6 (a, c) or DBA/2 (b) mice that had been immunized with ovalbumin in the presence (a, b) or absence (c) of syngeneic feeders were tested for their lytic activity on EL4 (C57BL/6, ●○▽▽) and P815 (DBA/2, ▲△) tumour targets in the presence (solid symbols) or absence (open symbols) of ovalbumin fragmented by trypsin digestion (a, b) or treated with CNBr (c).

Methods. Six-week-old C57BL/6 and DBA/2 (IFFA Credo) mice were intravenously injected either with a suspension of 2×10^6 syngeneic irradiated (2,000 rad) spleen cells in 0.5 ml of a solution of 4 mg ml⁻¹ ovalbumin (Sigma) phosphate-buffered saline (PBS) after 30 min incubation on ice or with 0.5 ml 0.2 mg ml⁻¹ ovalbumin in PBS. After 7 days animals were killed, spleens removed and cultured in single-cell suspensions in flat-bottomed 24-well plates (Costar) at 1×10^7 cells ml⁻¹ in IMDM (Gibco) supplemented with 10% fetal calf serum (Boehringer), 2 mM glutamine, non-essential amino acids, 1 mM hydroxypyruvate, penicillin at 100 international units ml⁻¹, streptomycin at 100 µg ml⁻¹, gentamycin at 50 µg ml⁻¹ (Sigma). Spleen cells obtained from animals immunized with feeders were grown in the presence of soluble ovalbumin at 0.7 mg ml⁻¹, to the other cultures CNBr-fragmented ovalbumin was added at 0.1 mg ml⁻¹. Ovalbumin was digested by trypsin (Sigma) as described previously¹⁶ and fragmented by CNBr (Sigma) according to ref. 17 (tOva and cOva, respectively). Both preparations were checked for complete degradation by gel filtration. For the CML assay, ⁵¹Cr-labelled target cells were preincubated with or without fragmented ovalbumin at 0.3 mg ml⁻¹ in supplemented IMDM for 60 min at 37 °C. For the actual test 1×10^4 target cells suspended in the same medium were incubated for 4½ h in round-bottomed microtitre plates (total volume, 0.2 ml; Costar) at effector to target ratios ranging from 30:1 to 1:1. Per cent specific lysis of these target cells was calculated as described elsewhere⁵.

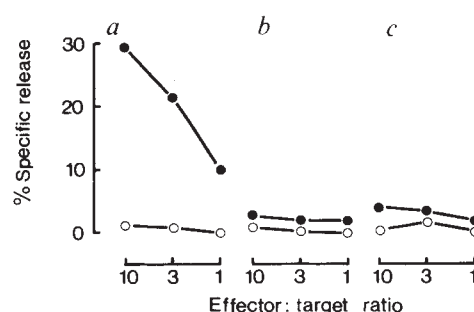


Fig. 2 Specificity of a cloned cytotoxic T lymphocyte line specific for CNBr-fragmented ovalbumin. The killing activity of C11 was tested against EL4 (a), P815 (b) and S.AKR (AKR/J, c) in the presence (●) and absence (○) of CNBr-fragmented ovalbumin. **Methods.** The CTL clone C11 was isolated by limiting dilution from a secondary culture of spleen cells of a C57BL/6 mouse immunized with ovalbumin, and maintained by weekly subculture with 0.1 mg ml⁻¹ CNBr-fragmented ovalbumin and irradiated syngeneic spleen cells in supplemented IMDM plus an exogenous source of IL-2. The CML assay was performed as for Fig. 1.

be observed. Preliminary experiments showed that in some cases minor but significant reactivity against ovalbumin could be obtained without prior *in vivo* priming (data not shown) suggesting that this 'background' reactivity might actually be directed against a component of the culture medium that can be presented by the target cell. The presence of syngeneic feeder cells at the time of immunization had no effect. Fragments of other proteins, for example pigeon cytochrome c or hen egg lysozyme (HEL), could not substitute for cOva or tOva. But CNBr-fragmented HEL could reveal specific MHC class I-restricted lytic activity after immunization with the soluble protein. Similar results could be obtained with blast cells as targets excluding any peculiar properties of the tumour cells (data not shown).

To further characterize the cell population responsible for this reactivity we established a cloned T-lymphocyte line, C11 (C57BL/6 derived) from a secondary *in vitro* culture directed against cOva. As determined by indirect surface immunofluorescence, C11 has the phenotype of a conventional CTL: CD1⁺, CD4⁻, CD8⁺ (data not shown). C11 had a restriction pattern identical to the uncloned population (Fig. 2). The syngeneic target EL4 was susceptible to lysis only in the presence of cOva, otherwise it was not killed by C11, nor were allogeneic targets. These findings indicate that the induced MHC class I-restricted response can be carried by conventional CD8⁺ CTL.

We wanted to examine whether this immune response against ovalbumin is directed against epitopes on this molecule that were expressed not only by chemical and enzymatic fragmentation, but also by intracellular protein degradation within the target cell. EL4 cells were transfected with BOVAnco expressing ovalbumin cDNA. The cell line EL4.Ova4-17 was subcloned and tested for susceptibility for lysis mediated by C11 (Fig. 3). The extent of lysis of the parental cell line in the presence of cOva (Fig. 3a) is comparable to that of the transfected cell line (Fig. 3b). In contrast, C11 could not kill EL4 grown in medium containing soluble undigested ovalbumin, excluding the possible explanation that secreted ovalbumin was taken up by EL4 and then presented in conjunction with MHC. It is noteworthy, however, that trypsin digestion of ovalbumin seems to destroy the epitope recognized by C11.

Although it cannot formally be excluded that ovalbumin is fragmented outside cells during the *in vivo* immunization, this explanation seems rather unlikely in view of findings that immunization with tOva, in all cases, failed to induce a measurable response (data not shown), whereas immunization with undigested ovalbumin always resulted in responses to epitopes

is suggested by several experimental lines of evidence. For instance, female mice reject syngeneic male skin-grafts after priming with male cells of another haplotype, and cytotoxic T lymphocytes of a given H-2 (mouse MHC) haplotype can be immunized against an array of minor and other antigens carried on cells of another haplotype^{12,13}. Therefore, it was proposed that foreign antigens arising from the cellular debris are taken up by host cells and presented in conjunction with class I MHC⁸.

We immunized mice with soluble ovalbumin, previously defined as a class II-restricted antigen. Irradiated syngeneic spleen cells incubated with a solution of ovalbumin were injected intravenously. Seven days later spleen cells were cultured in the presence of undigested ovalbumin for 5 days. Alternatively, soluble ovalbumin was only used for *in vivo* priming, and the spleen cells were grown *in vitro* together with CNBr-fragmented ovalbumin (cOva). To test for activity restricted by MHC class I molecules, tumour targets that do not express class II molecules were used (EL4, P815 and S.AKR). As demonstrated in Fig. 1, neither immunization protocol resulted in appreciable lytic activity towards allogeneic targets in the presence or absence of ovalbumin preparations. But C57BL/6 (H-2^b) and DBA/2 (H-2^d) mice immunized with ovalbumin mounted a specific response to the syngeneic targets, EL4 (H-2^b) or P815 (H-2^d) respectively, in the presence of their specific antigen, whereas in the absence of antigen only minor background activity could

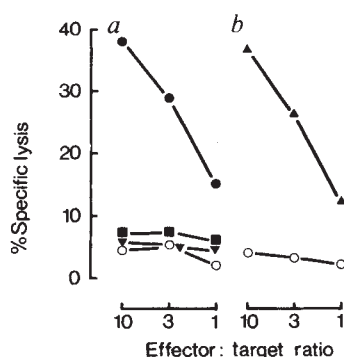


Fig. 3 Ovalbumin-transfected target cells can be lysed by the cloned CTL C11. The susceptibility to lysis of the parental line EL4 (○, ●, ■, ▼) in the presence of CNBr-fragmented (●), trypsin-digested (■) whole ovalbumin (▼) or uncoated (○) was compared to the ovalbumin transfected line EL4.Ova4-17 (△). **Methods.** Ovalbumin cDNA¹⁸ was introduced into the BM1Gneo expression vectors (H.K. & F. Melchers, in preparation) and EL4 cells were transfected with the resulting construct (BOVA neo). Neomycin-resistant EL4 clones were tested for ovalbumin expression by intracellular indirect immunofluorescence with a monoclonal mouse anti-ovalbumin antibody provided by D. Gray (data not shown). The line EL4.Ova4-17 was picked, subcloned by limiting dilution and tested for susceptibility for lysis mediated by C11. To test for presentation of undigested ovalbumin EL4 was grown in medium containing 10 mg ml⁻¹ ovalbumin for 18 h.

exposed by trypsin digestion. As it is clear from our data that at least part of the activity induced by our immunization protocols is carried by conventional CTL recognizing epitopes found on ovalbumin, there cannot be a complete separation of the two pathways of antigen presentation for MHC class I-restricted and MHC class II-restricted antigens. Our data indicate the existence of a mechanism allowing exogenous antigens to be presented in conjunction with MHC class I molecules in a similar way to antigens expressed within the cell. The fact that EL4 grown in soluble ovalbumin cannot be lysed by C11, in contrast to its counterpart transfected with the ovalbumin gene, points to the existence of a specialized antigen-presenting cell responsible for the induction of CTL. But, in contrast to previous proposals⁸, this antigen-presenting cell can take up soluble material. It should be of great interest to define the mechanisms and cell populations involved in the induction of CTL against soluble antigens further, as these cells might interfere with the immune response in the sense of a suppressive system.

We thank Drs H.-G. Rammensee, Gitta Stockinger and Craig C. Garner for critical reading of the manuscript. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche, Ltd. Co., Basel, Switzerland.

Received 11 August; accepted 13 August 1987.

1. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701-702 (1974).
2. Bevan, M. J. *J. exp. Med.* **142**, 1349-1364 (1975).
3. Katz, D. H., Hamaoka, T. & Benacerraf, B. *J. exp. Med.* **137**, 1405-1418 (1973).
4. Schwartz, R. A. *Rev. Immun.* **3**, 237-261 (1985).
5. Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457-467 (1985).
6. Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
7. Germain, R. N. *Nature* **322**, 687-689 (1986).
8. Bevan, M. J. *Nature* **325**, 192-194 (1987).
9. Townsend, A. R. M., Bastin, J., Gould, K. & Brownlee, G. G. *Nature* **324**, 575-577 (1986).
10. Maryanski, J. L., Pala, P., Corradin, G., Jordan, B. R. & Cerottini, J. C. *Nature* **324**, 578-579 (1986).
11. Parham, P. *et al. Nature* **325**, 625-628 (1987).
12. Fink, P. J., Weissman, I. L. & Bevan, M. J. *J. exp. Med.* **157**, 141-154 (1983).
13. Gooding, L. R. & Edwards, C. B. *J. Immun.* **124**, 1258-1262 (1980).
14. Shimonkevitz, R., Colon, S., Kappler, J. W., Marrack, P. & Grey, H. *J. Immun.* **133**, 2067-2074 (1984).
15. Bus, S., Sette, A., Colon, S. M., Miles, C. & Grey, H. M. *Science* **235**, 1353-1358 (1987).
16. Shimonkevitz, R., Kappler, J. W., Marrack, P. & Grey, H. *J. exp. Med.* **158**, 303-316 (1983).
17. Gross, E. & Witkop, B. *J. biol. Chem.* **237**, 1856-1860 (1962).
18. McReynolds, L. *et al. Nature* **273**, 723-728 (1978).

Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer

S. L. Naylor*, B. E. Johnson†, J. D. Minna† & A. Y. Sakaguchi*

* Department of Cellular and Structural Biology, The University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, Texas 78284, USA

† NCI-Navy Medical Oncology Branch, National Cancer Institute, Naval Hospital and Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814, USA

Specific chromosomal deletions sometimes associated with tumours such as retinoblastoma (chromosome 13q14)¹ and Wilm's tumour (chromosome 11p13)² have led to the hypothesis that recessive genes may be involved in tumorigenesis³. This hypothesis is supported by demonstration of allele loss specific for these regions using polymorphic DNA markers⁴⁻⁹ and by the isolation of a complementary DNA clone for the retinoblastoma gene¹⁰. A cytogenetic deletion in chromosome 3 (p14-p23) was reported in small-cell lung cancer (SCLC) by Whang-Peng *et al.*^{11,12}. At least one homologue of chromosome 3 was affected in the majority of SCLC tumours; however, the multiple chromosomal changes seen presented the possibility that chromosome 3 was rearranged, not deleted. We used polymorphic DNA probes for chromosome 3p and compared tumour and constitutional genotypes of nine SCLC patients. Our data show loss of alleles of chromosome 3p markers in tumour DNA of all nine patients supporting the hypothesis that this region contributes to tumorigenesis in SCLC.

DNA samples prepared from tumour and normal tissues of the same individual were obtained from cultured cell lines H128

Table 1 Chromosome 3 markers in SCLC

	DIS1 pH3B2 HindIII	D3S2 p12-32 HpaI	D3S3 pMS1-37 HpaI	D3S1 HS-3 HindIII	GP PstI	SST EcoRI
Group 1:						
H209BL	1,2	1,2	1,2	1,1	-	1,2
H209	1	1	1	1,1	-	1,2
N-1	1,2	1,1	1,1	1,2	1,2	1,1
SCLC-1	2	1,1	1,1	1,2	1,2	1,1
N-2	2,2	1,2	1,1	1,2	1,2	1,2
SCLC-2	2	2	1,1	1,2	1,2	1,2
N-3	1,2	1,1	1,2	1,2	1,2	1,1
SCLC-3	2	1	1	1,2	1,2	1,1
N-5	1,2	1,2	1,2	1,1	1,2	1,2
SCLC-5	2	1	1	1,1	1,2	1,2
N-6	2,2	1,2	1,2	1,2	1,1	1,2
SCLC-6	2	1	2	1,2	1,1	1,2
N-7	1,1	1,2	1,1	1,2	1,1	1,1
SCLC-7	1	2	1,1	1,2	1,1	1,1
Group 2:						
H128BL	2,2	1,2	1,2	1,2	-	1,1
H128	2	1	1	2	-	1
N-4	1,2	1,2	1,2	1,2	1,1	1,1
SCLC-4	1	2	1	1	1	1
Other lung cancers:						
N	2,2	1,2	1,1	1,2	2,2	1,1
Large cell	2,2	1,2	1,1	1,2	2,2	1,1
N	1,2	2,2	1,1	1,1	1,2	1,1
Adenocarcinoma	1,2	2,2	1,1	1,1	1,2	1,1

The numbers 1 and 2 refer to the alleles of a given locus. The copy number of homozygous markers is not known in the tumour samples. A dash indicates that the sample was not scored for a given polymorphism. Transferrin and somatostatin were not informative with *EcoRI* and *BamHI*, respectively, for any of these individuals nor was pH3E4 for *HindIII* digests. Locus *DIS1* is now called *DNF15S1*.

and H209 with matching B-cell lines (H128BL and H209BL) and seven autopsy samples of normal and tumour tissue. All cases were histologically confirmed as small-cell lung cancer. Two non-small-cell lung cancers were examined which included a large-cell tumour and an adenocarcinoma of the lung. DNA samples were digested with the appropriate restriction endonucleases to detect DNA polymorphisms.

Eight probes detecting DNA polymorphisms localized to chromosome 3 were used to study the individual chromosome 3 homologues in SCLC tumours and normal tissue from each patient. Two probes lie within 3p14-p23, the region commonly deleted in SCLC as determined by karyotypic analysis. Probe pMS1-37 (detecting the locus *D3S3*) is localized to chromosome 3 (ref. 13) band p14.2 (ref. 14) and detects an *MspI* polymorphism¹⁵, whereas probe p12-32 (*D3S2*) is located at band p21 and has 2 alleles in *MspI* digests¹⁶. The probe λ H3(*D1S1*, *DNF15S1*)¹⁷ has recently been reported to detect sequences on chromosome 3 as well as on chromosome 1 (refs. 18, 19), and the data of Donlon and Magenis²⁰ suggest that this marker localizes to 3p21. Two subclones of λ H3(*D1S1*), pH3H2 (a 2-kilobase (kb) *HindIII* fragment) and pH3E4 (a 4-kb *EcoRI* fragment) were used to detect two independent *HindIII* polymorphisms which have been shown to be located on chromosome 3 (refs. 18, 19). The DNA markers on the long arm of chromosome 3 are HS-3 (*D3S1*) detecting a *HindIII* polymorphism at q12; transferrin (*TF*), an *EcoRI* polymorphism at q21;

ceruloplasmin (*CP*) a *PstI* polymorphism at q25; and somatostatin (*SST*) at q28 with polymorphisms detected with *BamHI* and *EcoRI*.

The chromosome 3p markers *D3S2*, *D3S3* and *D1S1* were informative (that is, two different alleles were present in normal tissue) in all nine individuals, with six individuals heterozygous for two or more markers and three heterozygous for only one marker. Six individuals were heterozygous for *D3S3* alleles in DNA from normal tissue; all six had only one of the two alleles in the DNA from tumour tissue (see Fig. 1 and Table 1). Seven individuals were heterozygous for *D3S2* alleles in normal tissues, and had lost a *D3S2* allele in DNA from all of the corresponding tumours (see Fig. 1 and Table 1). All five individuals heterozygous for *D1S1* in normal tissue showed only one allele in the tumour. There does not appear to be a selective loss of a particular allele for any of the markers as either allele can be retained in different tumours. Tumour cell line H209, started from an untreated patient, had lost alleles of both *D3S3* and *D3S2* on chromosome 3p whereas markers on other chromosomes did not show allele loss.

Examination of polymorphisms over the length of chromosome 3 gave insight into the evolution of the chromosome 3 aberrations in SCLC. Table 1 lists the results for nine SCLC cases and two cases of large cell and adenocarcinoma of the lung examined with each of the chromosome 3 DNA probes. The samples fell into two groups: in the first and largest group

Table 2 Densitometry of DNA samples

		3p/3q				3p/3q			
		D3S3 4.8 kb allele	T/N	D3S3 3.6 kb allele	T/N	D3S2 2.9 kb allele	T/N	D3S2 1.3 kb allele	T/N
		D3S1		D3S1		D3S1		D3S1	
SCLC-2	Tumour	0		0.39		0		0.34	
	Normal	0		0.42	0.93	0.55	0	0.43	0.80
SCLC-3	Tumour	0		1.02		0.75		0	
	Normal	1.02	0	0.69	0.68	1.04	0.72	0	
SCLC-4	Tumour	0		0.54		0		0.08	
	Normal	1.14	0	3.84	0.14	0.80	0	0.40	0.20
SCLC-5	Tumour	0		0.64		0.45		0	
	Normal	1.32	0	2.19	0.29	0.94	0.48	0.68	0
SCLC-6	Tumour	0.95		0		0.87		0	
	Normal	0.72	1.31	0.40	0	0.73	1.19	0.70	0
Large cell	Tumour	0		1.86		0.67		0.90	
	Normal	0		1.73	1.08	1.07	0.62	0.62	1.45
Adeno	Tumour	0		2.24		0		2.56	
	Normal	0		2.54	0.88	0		1.82	1.41

		3p/15				3p/15			
		D3S3 4.8 kb allele	T/N	D3S3 3.6 kb allele	T/N	D3S2 2.9 kb allele	T/N	D3S2 1.3 kb allele	T/N
		D15S1		D15S1		D15S1		D15S1	
SCLC-2	Tumour	0		0.66		0		0.19	
	Normal	0		0.74	0.89	0.50	0	0.38	0.50
SCLC-3	Tumour	0.32		0.80		0.86		0	
	Normal	0.80	0.40	0.33	2.43	0.65	1.33	0	
SCLC-4	Tumour	0		0.88		0		0.22	
	Normal	0.45	0	1.52	0.58	0.42	0	0.22	1.00
SCLC-5	Tumour	0		0.65		0.80		0	
	Normal	0.83	0	1.38	0.47	0.95	0.84	0.69	0
SCLC-6	Tumour	1.92		0		3.32		0	
	Normal	1.47	1.31	0.82	0	1.97	1.68	1.89	0
Large cell	Tumour	0		1.90		1.49		2.02	
	Normal	0		2.72	0.70	2.69	0.55	1.55	1.30
Adeno	Tumour	0		2.71		0		4.01	
	Normal	0		2.29	1.18	0		5.20	0.77

To determine whether allele loss was associated with duplication of the remaining allele, scanning densitometry was performed on autoradiographs. To eliminate some of the variables in this method, the chromosome 3p data were normalized to probes on chromosome 3q and other chromosomes. Southern blots, which had been hybridized to chromosome 3p probes, were stripped of probe and hybridized to other probes. The 3q probe used was HS-3 (*D3S1*). The control chromosome used was pMS1-14 (*D15S1*). Similar results were obtained with p1E8 (*D13S4*) (data not shown). The autoradiographs were examined with a Hoefer scanning densitometer (GS300).

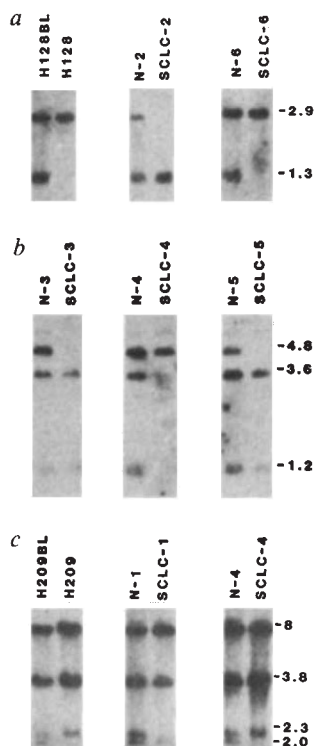


Fig. 1 Loss of alleles on chromosome 3p in small-cell lung cancer. Pairs of DNA samples from normal (left of each pair) and tumour (right) tissue from the same individual. *a*, *Msp*I digests hybridized to probe p12-32 (*D3S2*); allele 1 is 2.9 kb and allele 2 is 1.3 kb. *b*, *Msp*I digests hybridized to probe pMS1-37 (*D3S3*); allele 1 is 3.6+1.2 kb and allele 2 is 4.8 kb. *c*, *Hind*III digests hybridized to probe pH3H2; allele 1 is 2.3 kb and allele 2 is 2.0 kb. The 8 and 3.8 kb fragments are derived from both chromosomes 1 and 3.

Methods. DNA from each sample (5 µg) was digested and hybridized as described previously²¹. Filters were washed and exposed to X-ray films at -70 °C using intensifying screens for 1-3 days.

of tumours, loss of alleles was detected only for markers on the short arm of chromosome 3. These data are consistent with an interstitial or terminal deletion in chromosome 3 or perhaps somatic recombination. The karyotype of tumour cell line H209 (ref. 11) demonstrates the presence of a terminal deletion of chromosome 3. Densitometry of X-ray films indicated that chromosome 3 was probably hemizygous for short-arm markers in SCLC-2, SCLC-3, SCLC-5 and SCLC-6 (Table 2). Because of the aneuploidy of these tumours, however, densitometry was not always conclusive.

Chromosomal nondisjunction, as well as somatic recombination, are postulated to have occurred in chromosome 13 in certain instances of retinoblastoma⁴. Group 2 consists of two sample pairs (H128 and SCLC-4) in which heterozygous markers seen in both the short and long arm of chromosome 3 in DNA from normal tissue are reduced to single alleles in DNA from the tumours. Densitometry of SCLC-4 (Table 2) suggests that chromosome 3 is hemizygous in the tumour. The karyotyped tumour cell line H128 has several copies of intact as well as deleted chromosomes 3 (ref. 11), all having the same set of alleles remaining for *D3S1*, *D3S2*, and *D3S3* (Table 1).

Because SCLC is highly aneuploid^{11,12} we tested DNA markers on several other chromosomes (5, 8, 9, 10, 12, 13, 14, 15, 20, 21, 22). Some tumours showed loss of alleles on chromosomes 5, 13, 14, 15, 20 and 21. Tumours 1-6 were from patients who had received chemotherapy and four showed loss of an allele of one or more markers. Individual 7 had not received chemotherapy and all markers tested (other than chromosome 3) appeared to be the same in the tumour and in the normal tissue.

Karyotypic studies of SCLC using both cultured cell lines and direct tissue preparations have yielded variable results. Whang-Peng *et al.*^{11,12} found the deletion in 3p in all 16 SCLC cases analysed whether or not the individual had been treated with chemotherapy. In contrast, Wurster-Hill *et al.*²¹ found a chromosome 3 deletion in only 3 out of 15 cases studied. Other groups, however, have reported data similar to Whang-Peng: Yunis²² found four cases of deletion in four SCLC tumours, Falor *et al.*²³ reported deletions in three of three tumours, Buys *et al.*²⁴ found deletions in three of three and Ibson *et al.*³¹ eight of twelve tumours. Our studies using molecular probes demonstrate loss of chromosome 3 alleles at high frequency in SCLC (nine of nine cases). The two tumour cell lines and all seven of the SCLC samples obtained directly from the patient without intervening cell culture show loss of heterozygosity for 3p markers. Allele loss would not necessarily be observable karyotypically, perhaps accounting for the individuals which do not show a cytogenetic deletion in chromosome 3p in previous studies.

All karyotypic studies of SCLC have revealed a large number of chromosomal aberrations in the tumours^{11,12,21-24}. Our studies with DNA markers for other chromosomes confirm the presence of a number of chromosomal changes in SCLC (data not shown). In contrast, studies with embryonal and childhood tumours⁴⁻⁹ and acoustic neuroma²⁵ have shown little involvement of chromosomes other than the primary lesion. In adult bladder cancer Fearon *et al.*²⁶ also noted loss of alleles on a few chromosomes other than the commonly lost chromosome 11 markers.

The p21 region of chromosome 3 has been associated with at least two other types of adult cancer: a 3;8 translocation²⁷ and a 3;11 translocation²⁸ in renal cell carcinoma and chromosome 3p deletions and translocations in melanomas²⁹. Hansen *et al.*³⁰ have given evidence that osteosarcomas that develop later in patients surviving retinoblastoma also involve loss of alleles on chromosome 13. These data imply that similar genetic loci may be involved in histologically dissimilar tumours. Whether the same chromosome 3 loci are involved in SCLC, renal cell carcinoma and melanoma remains to be determined. Analysis of DNA markers in these tumours will test the hypothesis that these tumours share a common chromosomal mechanism for tumorigenesis.

We thank Stella Notzon, Lisa McDonald, Julie Gardner, Scott Marshall and Amy Solomon for technical assistance, and Diana Hottle for preparation of the manuscript. This work was supported by the Council for Tobacco Research and the National Cancer Institute. The probes used in this study were generous gifts of Ray White (pMS1-37 and p12-32), Mary Harper and Grady Saunders (λH3), James Gusella (HS-3), Funmei Yang and Barbara Bowman (TF and CP), Graeme Bell (SST) and A. Gazdar for cell lines and review of pathology. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Navy or the Department of Defense.

Received 24 May; accepted 10 July 1987.

1. Francke, U. *Cytogenet. Cell Genet.* **16**, 131-134 (1976).
2. Riccardi, V. M., Sujansky, E., Smith, A. C. & Francke, U. *Pediatrics* **61**, 604-610 (1978).
3. Knudson, A. G. A. *Rev. Genet.* **20**, 231-251 (1986).
4. Cavenee, W. K. *et al. Nature* **305**, 779-784 (1983).
5. Koufos, A. *et al. Nature* **309**, 170-172 (1984).
6. Orkin, S. H., Goldman, D. S. & Sallan, S. E. *Nature* **309**, 172-174 (1984).
7. Reeve, A. E. *et al. Nature* **309**, 174-176 (1984).
8. Fearon, E. R., Vogelstein, B. & Feinberg, A. P. *Nature* **309**, 176-178 (1984).
9. Glaser, T. *et al. Nature* **321**, 882-887 (1986).
10. Friend, S. H. *et al. Nature* **323**, 643-646 (1986).
11. Whang-Peng, J. *et al. Cancer Genet. Cytogenet.* **6**, 119-134 (1982).
12. Whang-Peng, J. *et al. Science* **215**, 181-182 (1982).
13. DeMartinville, B., Shafer, M., White, R. & Francke, U. *Cancer Genet. Cytogenet.* **37**, 531-532 (1984).
14. Gerber, M. J., Miller, Y. E., Drabkin, H. A. & Scoggins, C. H. *Cytogenet. Cell Genet.* **42**, 72-74 (1986).
15. Barker, D., Schafer, M. & White, R. *Cell* **36**, 131-138 (1984).
16. Naylor, S. L., Sakaguchi, A. Y., Barker, D., White, R. & Shows, T. B. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2447-2451 (1984).
17. Harper, M. E. & Saunders, G. F. *Chromosoma* **83**, 431-439 (1981).
18. Goode, M. E., vanTuinen, P., Ledbetter, D. H. & Daiger, S. P. *Am. J. hum. Genet.* **38**, 437-446 (1986).

19. Carritt, B., Welch, H. M. & Parry-Jones, N. J. *Am. J. hum. Genet.* **38**, 428-436 (1986).
20. Donlon, T. A. & Magenis, R. E. *Cytogenet. Cell Genet.* **37**, 454-455 (1984).
21. Wurster-Hill, D. H. *et al. Cancer Genet. Cytogenet.* **13**, 303-340 (1984).
22. Yunis, J. J. *Science* **221**, 227-236 (1983).
23. Falor, W. H., Ward-Skinner, R. & Wegryn, S. *Cancer Genet. Cytogenet.* **16**, 175-177 (1985).
24. Buys, C. H. C. M., Van der Veen, A. J. & deZey, L. *8th International Chromosome Conference, Lubeck GFR*, 1983.
25. Seizinger, B. R., Martuza, R. L. & Gusella, J. F. *Nature* **322**, 644-647 (1986).
26. Fearon, E. R., Feinberg, A. P., Hamilton, S. H. & Vogelstein, B. *Nature* **318**, 377-380 (1985).
27. Cohen, A. J. *et al. New Engl. J. Med.* **301**, 592-595 (1979).
28. Pathak, S., Strong, L. C., Ferrell, R. E. & Trindade, A. *Science* **217**, 939-941 (1982).
29. Pederson, M., Strand, R. & Wang, N. *Am. J. hum. Genet.* **36**, 1095 (1984).
30. Hansen, M. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6216-6220 (1985).
31. Ibsen, J. M. *et al. J. Cell Biochem.* **33**, 267-288 (1987).

Maternal inhibition of hepatitis B surface antigen gene expression in transgenic mice correlates with *de novo* methylation

Michelle Hadchouel*, Hend Farza*, Dominique Simon†, Pierre Tiollais* & Christine Pourcel*‡

* Recombinaison et Expression Génétique, INSERM U163, CNRS UA 271; † Génétique des Mammifères, Institut Pasteur, 25 rue de Dr Roux, 75724 Paris, Cedex 15, France

Differential modifications of the genome during gametogenesis result in a functional difference between the paternal and maternal genomes at the moment of fertilization¹⁻⁴. A possible cause of this imprinting is the methylation of DNA^{5,6}. The insertion of foreign DNA into transgenic mice allows the tagging of regions that are differentially methylated during gametogenesis. We describe here a transgenic mouse strain in which the expression of the hepatitis B surface antigen gene is irreversibly repressed following its passage through the female germ line. This inhibition is accompanied by the methylation of all the *HpaII* and *HhaI* sites within the foreign gene, which we have shown to be integrated into a site on chromosome 13. The irreversibility reported here contrasts with what is found with other transgenic mice sequences which are reversibly methylated after passage through the male or female germ line^{7,8}, though in both cases methylation appears to be important in the imprinting process.

We have recently described two transgenic male mice containing hepatitis B virus (HBV) sequences in their genome and expressing hepatitis B surface antigen (HBsAg)⁹. The sequences incorporated consisted of a 2.8-kilobase (kb) fragment of the HBV genome cloned into the *Bam*HI site of plasmid pBR322 (to give recombinant PAC2). Both mice, Tg(0HBV)E11-Pas (referred to as E11) and Tg(13HBV)E36-Pas (referred to as E36), contained one copy of the recombinant plasmid integrated close to the *Hind*III site used to linearize the plasmid (Fig. 1). The flanking sequences differed in the E11 and E36 strains, as judged by restriction mapping and hybridization with flanking sequences from the E36 strain.

Although the arrangement of the viral DNA was similar in both strains we observed two different types of transmission. The male mouse E11 transmitted the HBV-recombinant sequences to a male F₁ mouse produced by *in vitro* fertilization, then normally to next generation (N₂) males and females. In subsequent crosses, HBV DNA positive progeny always produced HBsAg irrespective of whether the plasmid was transmitted through the male or female N₂ mice (Fig. 2). The second mouse, E36, transmitted the HBV sequences without rearrangement to all its male and female F₁ progeny (designated H/+), and these always expressed HBsAg. HBV positive progeny from H/+ males in the second generation also always expressed HBsAg (Fig. 3a). Progeny of H/+ females receiving the HBV DNA did not, on the other hand, produce HBsAg and there was no transcription of the *S* gene in the liver. This was not due to gross

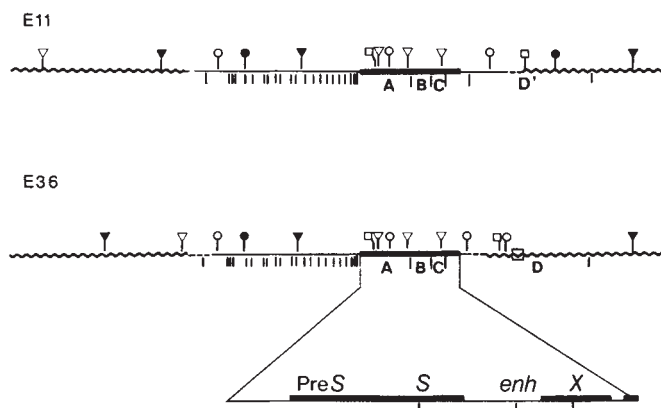


Fig. 1 Restriction map of the integrated PAC2 sequences in strains E11 and E36. A *Bgl*III HBV DNA fragment (solid line) was cloned into the *Bam*HI site of pBR322 (thin line). The HBV sequences are enlarged on the bottom line to show the genetic organization. *PreS* and *S* encode the HBsAg polypeptides, *enh* is an enhancer region and *X* is an unidentified gene. Approximately one copy of the recombinant plasmid is inserted into the mouse genome (wavy line). A restriction map was made using *Eco*RI (sites shown as open circles), *Bam*HI (open triangle), *Bgl*III (open square), *Pst*I (dark circle) and *Pvu*II (dark triangle). The bars below represent *Hpa*II/*Msp*I sites. The E36 insertion was cloned into the λ vector L 47-1 and a probe (box) in the right flanking mouse DNA was isolated. It detected *Bcl*I and *Taq*I polymorphisms that segregated with markers for chromosome 13 (refs 10 and 11). For designation of A, B, C, D and D', see text and Fig. 4.

rearrangement of the HBV sequences, as judged by restriction map analysis (data not shown). The mice were designated H^m/+ for modified or maternal HBV sequences. The loss of HBsAg expression was reproducibly observed when HBsAg-positive N₂ females were mated to normal males. The data show that female-transmitted HBV DNA sequences do not allow HBsAg expression in either the male or female progeny. To test the reversibility of this phenomenon, progeny from H^m/+ males and females were analysed for HBsAg expression. Among 17 HBV DNA-positive offspring from an H^m/+ male and 7 offspring from an H^m/+ female, none expressed HBsAg (Fig. 3a). This showed that the modification was not reversible even when the modified allele was transmitted by a male.

F₁ intercrosses were then set up to obtain homozygous offspring (Fig. 3b). The homozygotes received a modified allele of maternal origin and a normal one of paternal origin (H/H^m) which allowed HBsAg expression. HBsAg serum levels were similar in heterozygotes and homozygotes showing that neither H nor H^m was dominant. When crossed to normal mice H/H^m females produced 100% HBV positive, HBsAg negative progeny whereas H/H^m males gave 50% HBsAg positive animals and 50% HBsAg negative animals, confirming the irreversibility of the modification on passage through the male germ line. Homozygous H/H^m animals and their progeny developed normally, showing that insertion of the plasmid had not directly affected a region essential for either development or fertility. H^m/H^m homozygotes likewise developed normally and were fertile.

To test whether the lack of HBsAg gene expression in H^m/+ mice was due to a change in methylation status we carried out restriction enzyme digestions with *Msp*I and *Hpa*II. HBV sequences in the PAC2 plasmid contain three *Msp*I/*Hpa*II sites and pBR322 contains 22 *Msp*I/*Hpa*II sites (Fig. 1). Liver DNA from H/+ and H^m/+ mice was cleaved with *Msp*I and hybridized to an HBV probe (Fig. 4, lane 1). Three fragments of 1.3, 0.6 and 0.4 kb (bands A, B and C) produced by cleavage within the HBV DNA were detected, plus a 3.8 kb fragment (band D) corresponding to a site in the flanking DNA. A different pattern was obtained after digestion with *Hpa*II. In H/+ males and

‡ To whom correspondence should be addressed.

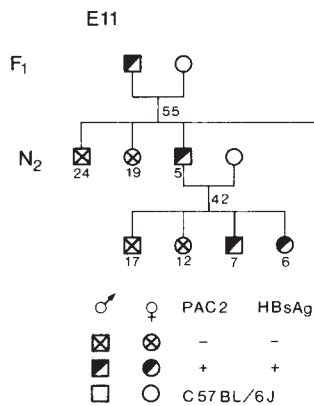


Fig. 2 Pedigree chart of strain E11. The founder male produced one offspring by *in vitro* fertilization. This mouse (male F₁) was crossed to C57BL/6-J-Pas females and produced male and female offspring, referred to as N₂. The mice were tested for HBV DNA using the dot-hybridization technique⁹. Serum HBsAg was measured with the AUSRIA radioimmunoassay (Abbott). The total number of offspring is indicated under each cross and the number of mice of each type indicated under the corresponding symbol.

females (Fig. 4, lanes 2 and 3), bands A, B and C were present without band D, showing that the first site in the cellular DNA was methylated. Additional bands corresponding to A+B and A+B+C arose from partial methylation of the HBV DNA. Hybridization with a pBR322 probe showed that the plasmid sequences were completely digested by *Hpa*II (data not shown). When liver DNA from male and female H^m/+ was analyzed in the same way the pattern obtained with *Hpa*II digestion was strikingly different (Fig. 4, lanes 4 and 5). Hybridization was found only in the high molecular weight DNA with both an HBV and a pBR322 probe, although complete digestion of the cellular DNA could be demonstrated by hybridization with a control probe. DNA from H^m/+ mice, but not from H/+ mice, was also resistant to *Hha*I digestion (data not shown). The same result was found with DNA extracted from several organs of a H^m/+ mouse, suggesting that the *de novo* methylation occurred in the early embryo or during oogenesis.

Liver DNA extracted from males and females of the E11 strain was also digested by *Msp*I (Fig. 4, lane 6) and *Hpa*II (Fig. 4, lanes 7 and 8) and hybridized to HBV. The A, B and C bands were detected in both *Msp*I and *Hpa*II digestions plus a 3 kb *Msp*I fragment (band D') corresponding to a site in the flanking sequence. An additional band corresponding to A+B was also present.

The HBV sequences integrated in the E36 strain plus 2 kb of mouse flanking sequences were cloned, and a genomic probe

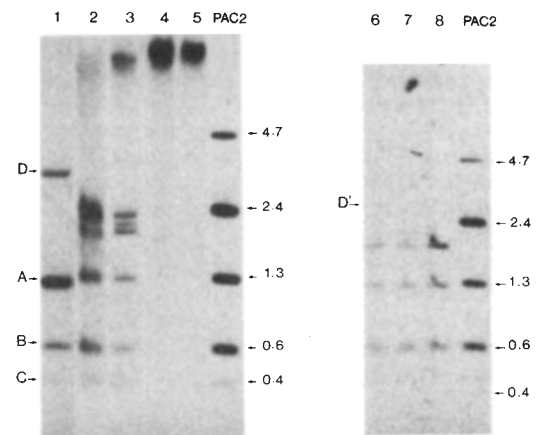
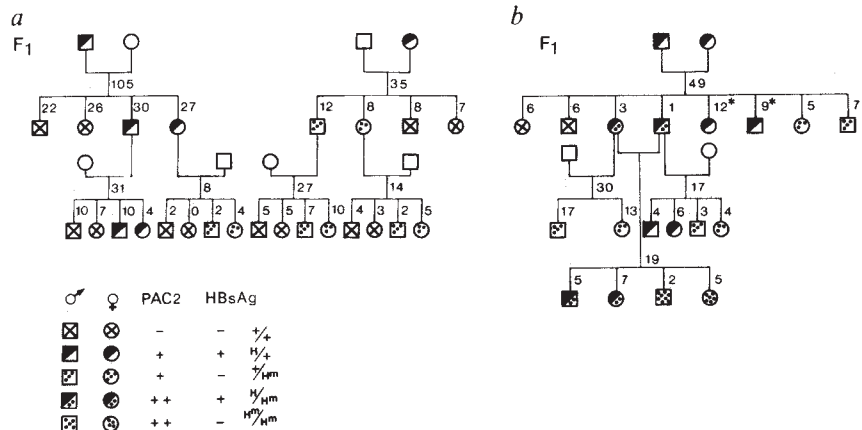


Fig. 4 Analysis of the methylation status of E11 and E36 liver DNA. DNA (15 µg) from the liver of males and females of the two strains were cut with *Msp*I or *Hpa*II and analyzed by electrophoresis on a 1.2% agarose gel and blot-hybridization⁹. Lanes 1 and 6: *Msp*I digestion of E36 and E11 liver DNA. Lanes 2 and 3: *Hpa*II digestion of male and female H^m/+E36 liver DNA. Lanes 4 and 5: *Hpa*II digestion of male and female H^m/+E36 liver DNA. Lanes 7 and 8: *Hpa*II digestion of male and female E11 liver DNA. The probe used to hybridize was the complete HBV DNA. PAC2 restricted with *Eco*RI or *Hpa*II (together in lane PAC2) served as a size marker. For designation of bands A, B, C, D and D', see text and Fig. 1.

devoid of repetitive DNA was isolated (Fig. 1). With this probe we have located the insertion site on chromosome 13 using an interspecific back cross between *Mus spretus* (SPE/Pas) and *Mus musculus domesticus* (C57BL/6-J-Pas)^{10,11}. This region is methylated over a large distance in different tissues (including sperm) from adult male and female mice, in 11 day-old embryos, and in the F9 cell line (data not shown). Oocyte DNA was not tested.

The *de novo* methylation of HBV sequences, on passage through the female germ line, is a consequence of its insertion into a particular site in the mouse genome and represents another example of a position effect¹². It could reflect differential methylation of this region but this would necessitate its demethylation at some point during gametogenesis. One possible interpretation is that demethylation occurs in oocytes, followed by *de novo* methylation of the female allele only, in the early embryo. This could be another manifestation of imprinting, although different from that described by Reik *et al.*⁷ and Sapienza *et al.*⁸. Chromosome 13 is not, however, thought to be involved in the parental effect because zygotes with maternal or paternal disomy survive normally^{13,14}.

Fig. 3 Pedigree chart of strain E36. The founder male produced two types of offspring containing HBV DNA, only one of which produced HBsAg⁺. Mice of this latter type were selected for further studies. The first generation is referred to as F₁. Progeny were tested for the presence of HBV DNA and serum HBsAg as described in the legend to Fig. 2. *a*, Crosses between heterozygotes for the HBV insertion (referred to as H/+) and normal C57BL/6-J-Pas mice (open symbols). Three different F₁ males or F₁ females were crossed with normal mice. The modified HBV insertion transmitted by females is referred to as H^m. *b*, Crosses between two F₁ heterozygotes. The progeny of three different pairs of heterozygotes is reported. Homozygosity was first detected by hybridization and confirmed by pedigree analysis. * Total includes undetected homozygotes. H/H^m homozygotes were crossed to produce mice with two modified HBV insertions H^m/H^m which were fertile when crossed with normal mice (not shown).



Our findings also show that methylation of HBV sequences prevents *S* gene expression. Elsewhere we show (manuscript in preparation), that an *HpaII* site in the *S* gene is specifically methylated in organs of H/+ mice that do not express HBsAg. This site is demethylated during liver differentiation. In the maternal inhibition described here, the extent of the methylation probably prevents reversibility. This can be compared with the irreversible inactivation of proviral sequences by methylation¹⁵.

We thank C. Babinet for his constant support and interest throughout this work and P. Avner, C. Cremisi, J. L. Guenet and G. Vergnaud for discussions and reading of the manuscript.

Received 22 June; accepted 18 August 1987.

1. Surani, M. A. H. & Barton, S. C. *Nature* **308**, 548-550 (1984).
2. McGrath, J. & Solter, D. *Cell* **37**, 179-183 (1984).
3. Cattanaach, B. M. & Kirk, M. *Nature* **315**, 496-498 (1985).
4. Surani, M. A. H., Barton, S. C. & Norris, M. L. *Nature* **326**, 395-397 (1987).
5. Surani, M. A. H., Barton, S. C. & Norris, M. L. *Cell* **45**, 127-136 (1986).
6. McGrath, J. & Solter, D. *Nature* **308**, 550-551 (1984).
7. Reik, W., Collick, A., Norris, M. L., Barton, S. C. & Surani, M. A. *Nature* **328**, 248-251 (1987).
8. Sapienza, C., Peterson, A. C., Rossant, J. & Balling, R. *Nature* **328**, 251-254 (1987).
9. Babinet, C., Farza, H., Morello, D., Hadchouel, M. & Pourcel, C. *Science* **230**, 1160-1163 (1985).
10. Robert, B. *et al.* *Nature* **314**, 181-183 (1985).
11. Guenet, J. L. *Curr. Topics microbiol. Immun.* **127**, 109-113 (1986).
12. Jaenisch, R. & Jähner, D. *Biochim. biophys. Acta* **782**, 1-9 (1984).
13. Lyon, M. F., Ward, H. C. & Simpson, G. M. *Genet. Res.* **29**, 83-92 (1976).
14. Lyon, M. F. in *Radiation-Induced Damage in Man* (ed. Ishihara, T.) 327-346 (Liss, New York, 1983).
15. Jaenisch, R. *et al.* *Cell* **24**, 519-529 (1981).

p53 and DNA polymerase α compete for binding to SV40 T antigen

J. V. Gannon & D. P. Lane

Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Potters Bar, Herts EN6 3LD, UK

The large T antigen (T) of simian virus 40 is a multifunctional protein required for both viral DNA replication and cellular transformation¹. T antigen forms specific protein complexes with the host protein p53 in both virus-infected and transformed cells²⁻⁴. p53 has recently been shown to be an oncogene⁵⁻⁷, but its normal function is not clear. We previously established a radioimmunoassay⁸ to study the newly described complex between T antigen and DNA polymerase α , and have noted a similarity between the antigenic changes induced in T by the binding of both p53 and polymerase^{9,10}. We now extend this analysis to a larger collection of anti-T antibodies and formally establish that p53 and DNA polymerase α can compete for binding to the SV40 T antigen. At a critical concentration of the three components it is possible to detect a trimeric complex of T, p53 and DNA polymerase α . Our observations have important implications for the control by these nuclear oncogenes of viral and cellular DNA synthesis and viral host range in both normal and transformed cells. We present a model for the action of p53 in growth control.

A quantitative radioimmunoassay for the complex between T antigen and DNA polymerase α was established by coating plastic microtitre wells with any one of a set of well-characterized anti-DNA polymerase α monoclonal antibodies¹¹. These wells were then incubated with extracts of virus-infected cells and probed for the presence of DNA polymerase α -T complexes with a range of radiolabelled anti-T monoclonal antibodies. Control wells were coated with monoclonal antibodies of irrelevant specificity, and total T antigen in the extract was measured using an anti-T coated plate (Fig. 1a). Only a small fraction of the T antigen present in the extract appeared to be complexed to DNA polymerase α . The availability of large libraries of monoclonal antibodies to T, recognizing discrete epitopes on the protein, allowed an examination of the epitopes displayed by T bound to DNA polymerase α . Some of the anti-T

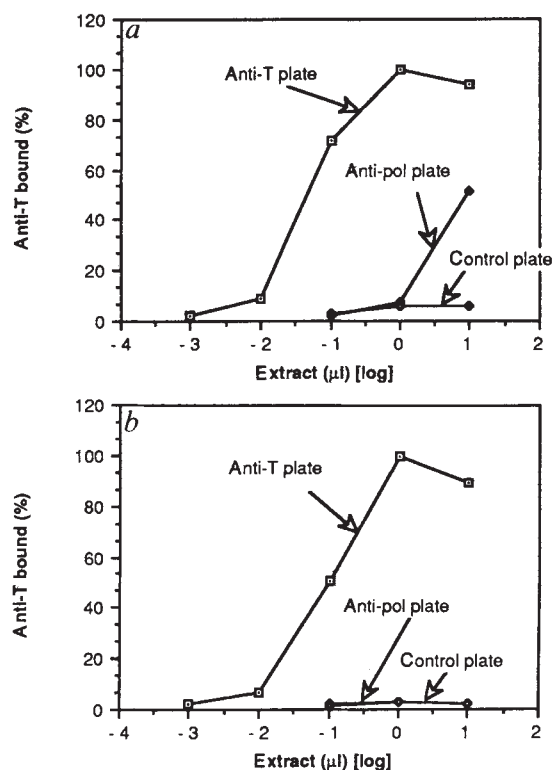


Fig. 1 Radioimmunoassay of the DNA polymerase α -T complex. T antigen and the DNA polymerase α -T complex present in an extract of Ad-5 SVR 111-infected 293 cells were estimated using a monoclonal antibody sandwich immunoassay with labelled anti-T antibodies and anti-T (Pab423, ref. 18) anti-DNA polymerase α (SJK 287-32, ref. 11), or control monoclonal solid phase antibodies. The anti-T label is a, Pab 419 (ref. 18) and b, Pab 250 (ref. 13). (A similar result to that shown in b was obtained using Pab 204 as label). Binding of each anti-T label is expressed as % of the maximum binding achieved for comparison between antibodies^{9,10}.

Methods. The cell extract was prepared from 293 cells grown on monolayers and infected with virus at a m.o.i. of 10 p.f.u. 24 h earlier¹⁴. The extract was prepared as an S100 (ref. 13) in 10 mM HEPES, pH 7.4, containing 1.5 mM $MgCl_2$ and 5 mM KCl (HYPO buffer) at a ratio of 1.5×10^8 cells ml^{-1} . All monoclonal antibodies were purified by protein A-Sepharose chromatography. For the solid phase, antibodies were used at a concentration of 30 $\mu g\ ml^{-1}$ to coat the wells of 96 well plastic microtitre plates. After coating overnight, plates were blocked by overnight incubation in a 20% solution of dried skimmed milk in PBS and then rinsed in PBS. Extracts were incubated on the antibody-coated plates overnight at 4°C and, after rinsing in PBS containing 0.1% NP40, probed with the appropriate iodinated antibody for 4 h before washing and counting the individual wells. All assay points were determined in duplicate. Antibodies were labelled with ¹²⁵I to specific activities of 5 $\mu Ci\ \mu g^{-1}$.

monoclonal antibodies, for example Pab 250 (binding site between amino acids 271 and 367 of T) and Pab 204 (binding site between amino acids 453 and 469 of T), could not detect the T present in complex with DNA polymerase α (Fig. 1b). These same antibodies are unable to detect T that is complexed to p53 (ref. 10). It appeared therefore that DNA polymerase α and p53 could be occupying similar sites on T antigen or inducing the same conformational changes in the protein.

To test this idea directly we established an *in vitro* assay for the assembly of the DNA polymerase α -T complex on the same principle as one used earlier for the T-p53 complex assembly¹⁰. The assay is capable of measuring the assembly of the T-DNA polymerase α complex, and both components of the complex can be independently titrated; a titration for T antigen is shown (Fig. 2). Maximum binding of the DNA polymerase α present

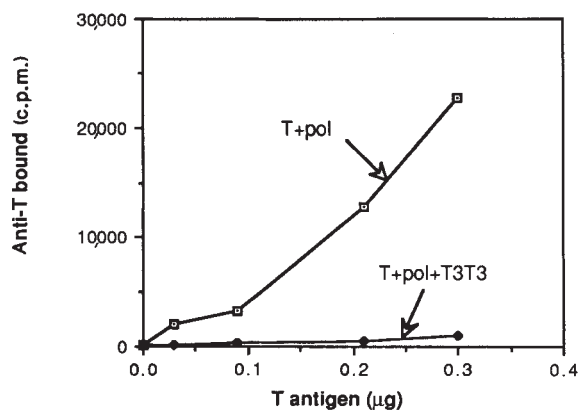


Fig. 2 Extracts of murine T3T3 cells block the *in vitro* binding of T antigen to DNA polymerase α . T antigen binding to DNA polymerase α *in vitro* was measured using an SJK 287-38 anti-DNA polymerase α solid phase and an iodinated anti-T, Pab 419, as a label. 50 μ l HeLa S100, a source of DNA polymerase α , was mixed with 50 μ l extract (10^8 cells ml^{-1} in 50 mM Tris, pH 8.0, containing 5 mM EDTA, 2 mM PMSF and 150 mM NaCl) of T3T3 cells, 'T+pol+T3T3', or 50 μ l buffer alone, 'T+pol'. Variable amounts of pure T antigen in 10 μ l buffer (20 mM HEPES, pH 7.4, 0.1 mM EDTA, 5 mM NaCl, 10% glycerol, 1% BSA) were added to the mix. After incubation for 2 h at 4 $^{\circ}\text{C}$ the mixtures were transferred to the solid phase antibody plate and incubated, washed, and probed with labelled antibody as described in Fig. 1.

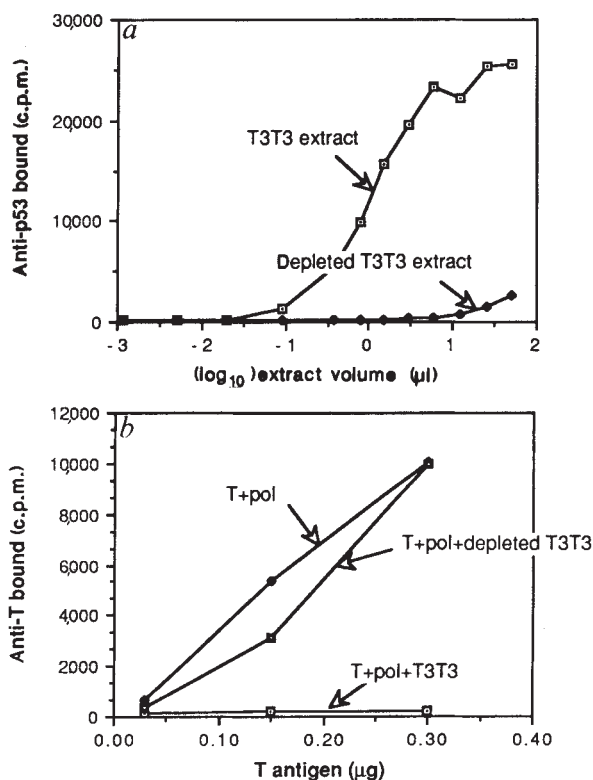


Fig. 3 Murine p53 blocks T antigen binding to DNA polymerase α . **a**, Immunoaffinity depletion of p53 from the T3T3 extract. An extract of T3T3 cells prepared as described in Fig. 2 was depleted of p53 by incubation with the anti-p53 antibody Pab 421 (ref. 18) covalently coupled to protein A-Sepharose beads¹⁴. The effectiveness of the depletion was monitored by titrating the starting extract and the depleted extract in a p53 sandwich immunoassay^{19,20}. This assay used the anti-p53 RA3 2C2 (ref. 21) as a solid phase and another anti-p53 200.47 (ref. 22) as the labelled probe. Wash conditions and incubation times are described in Fig. 1. **b**, T3T3 extracts depleted of p53 no longer inhibit the binding of T to DNA polymerase α . The assay was run as described in Fig. 2 except that, in the case of 'T+pol+depleted T3T3', the total T3T3 extract was replaced with 50 μ l T3T3 extract, depleted of p53 by immunoaffinity chromatography as shown in Fig. 3a.

in 50 μ l HeLa cell S100 extract^{12,13} is seen with 0.5 μ g of immunoaffinity purified T antigen¹⁴. From the conditions of the assay we can calculate that the minimum affinity constant of T for DNA polymerase α must be at least 10^6 M^{-1} and could greatly exceed that figure if only a subfraction of T is competent to bind DNA polymerase α . The subunit composition of the mammalian DNA polymerase α remains controversial¹⁵ and it is not known whether T interacts directly with the catalytic core polypeptide (against which the anti-DNA polymerase α antibodies used here are directed¹¹), or with one of the associated polypeptides. Indeed it is possible that T itself may alter the subunit composition of the holoenzyme, perhaps substituting for a cellular subunit. To ask whether p53 could inhibit the binding of DNA polymerase α to T, a competition experiment was set up in the *in vitro* assembly assay. An S100 extract of HeLa cells served as a source of DNA polymerase α and an extract of murine T3T3 cells (a line of spontaneously transformed 3T3 cells¹⁰ containing high levels of p53) was used as a source of p53; the two extracts were mixed and incubated with a range of T antigen concentrations. The T3T3 extract clearly inhibited the *in vitro* binding of DNA polymerase α to T (Fig. 2). To determine if p53 was the active component in the T3T3 extract responsible for this inhibition, the extract was selectively depleted of p53 by immunoabsorption on an anti-p53 monoclonal antibody column. A p53 radioimmunoassay confirmed that the extract was completely depleted of p53 (Fig. 3a). This depleted extract was unable to block the association of T and DNA polymerase α (Fig. 3b), confirming our hypothesis that p53 is the component responsible for the inhibition. Passing the T3T3 extract through control columns of anti-polymerase or anti-T antibodies did not remove p53 or deplete the extract of its inhibitory activity, though we could show these columns to be effective.

T antigen is found in infected and transformed cells in a variety of oligomeric forms¹. In particular the T-p53 complex is highly oligomeric, containing multiple T and p53 subunits^{1,3,4}. T purified by immunoaffinity chromatography¹⁴ is at least tetrameric, so that at subsaturating levels of p53 and DNA polymerase α it should be possible to assemble a complex of p53 and DNA polymerase α 'bridged' by T antigen. To identify such a complex, a variable amount of T3T3 extract was mixed

with pure T antigen after mixing with control buffer or with HeLa cell extract (Fig. 4). Using appropriate immunoassays, the formation of T-pol complex, (Fig. 4a), T-p53 complex (Fig. 4b), and the postulated pol-T-p53 complex (Fig. 4c) was measured. The results show (Fig. 4a) that detection of T-pol complex is, under these conditions, absolutely dependent on the presence of the HeLa extract and that the formation of this complex is suppressed by high concentrations of murine p53 (from the T3T3 extract). The formation of the T-p53 complex (Fig. 4b) is not dependent on, or suppressed by, the HeLa extract, but is dependent on the T3T3 extract. Finally, a trimeric complex of pol-T-p53 can be detected, using an anti-pol solid phase and an anti-p53 label (Fig. 4c). The formation of this complex reaches a sharp peak within the narrow T3T3 extract (p53) titration range and high concentrations of p53 inhibit the formation of the complex since all binding sites on the T oligomer are occupied by p53. The formation of this p53-DNA polymerase α complex only occurs in the presence of T antigen.

The finding that the cellular protein p53 can compete with DNA polymerase α for binding to T antigen is the first report of a direct effect of p53 on the biochemical activity of T antigen. If the formation of a complex between T antigen and DNA polymerase α is an essential step in SV40 DNA synthesis then p53 could potentially block this process. The p53 from non-permissive species binds more tightly to T compared to p53

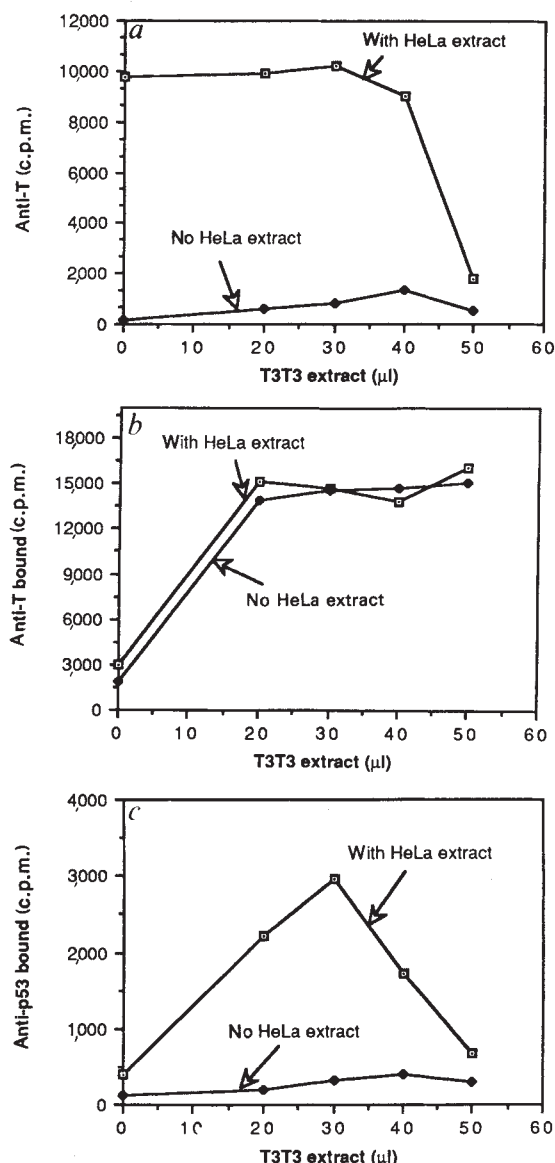


Fig. 4 Simultaneous analysis of the formation of T-pol, T-p53 and a trimeric pol-T-p53 complex. Extracts of T3T3 cells were mixed at different concentrations with extracts of HeLa cells or buffer alone and incubated with pure T antigen. Aliquots of the mixtures were assayed for the presence of T-pol complexes, T-p53 complexes and pol-T-p53 complexes. *a*, T-pol complex assay using an anti-T label (PAb 419) and an anti-DNA polymerase α (SJK 287-32) as solid phase. *b*, T-p53 complex assay using the same anti-T label as in *a* and an anti-p53 (RA3 2C2, ref. 21) as solid phase. *c*, pol-T-p53 complex assay using an anti-p53 (200.47, ref. 22) as label and an anti-DNA polymerase α (SJK 287-32) as solid phase.

Methods. Assay conditions and wash steps are described in Figs 1, 2 and 3. A range of volumes of T3T3 extract (0 μ l, 120 μ l, 180 μ l, 240 μ l and 300 μ l) prepared in HYPO buffer were adjusted to a final volume of 300 μ l in the same buffer. An equal volume of HeLa extract or buffer (no HeLa extract) was added and pure T antigen added to each mixture. After incubation two 100 μ l aliquots of each reaction mix were assayed in each of the three assays.

from permissive species¹⁶, and competition between p53 and DNA polymerase α for T could be a determinant of host permissiveness.

Indeed we recently learned that transfection of murine (but not human) p53 expression plasmids into COS cells leads to a dramatic reduction in the T-dependent replication of cotransfected SV40 origin containing plasmids. These results (J. Jenkins,

personal communication) provide a striking *in vivo* correlation to the *in vitro* studies presented here.

The anti-pol antibodies used in this study will, under certain assay conditions, react with murine DNA polymerase α . T will bind murine DNA polymerase α *in vitro*, provided the extracts have been depleted of murine p53, and as this association is also blocked by p53 (D.P.L. and J.V.G., manuscript in preparation), the T-pol interaction may be important in viral transformation as well as in lytic infection. The connection between p53 and DNA polymerase α resulting from their competitive binding to SV40 T antigen could provide insight into the normal cellular function of the p53 proto-oncogene. Possibly a cellular inhibitory factor that binds to DNA polymerase α might also bind p53. The increased amounts of p53 found as cells enter S phase¹⁷ could be necessary to remove this factor, and the high p53 concentrations found in a range of tumour cells might be causally linked to their uncontrolled DNA replication.

We thank Ed Harlow for anti-T antibodies, Yasha Gluzman for virus, Bruce Stillman for advice on replication extract preparation and John Jenkins for discussion of results before publication.

Received 29 May; accepted 20 August 1987.

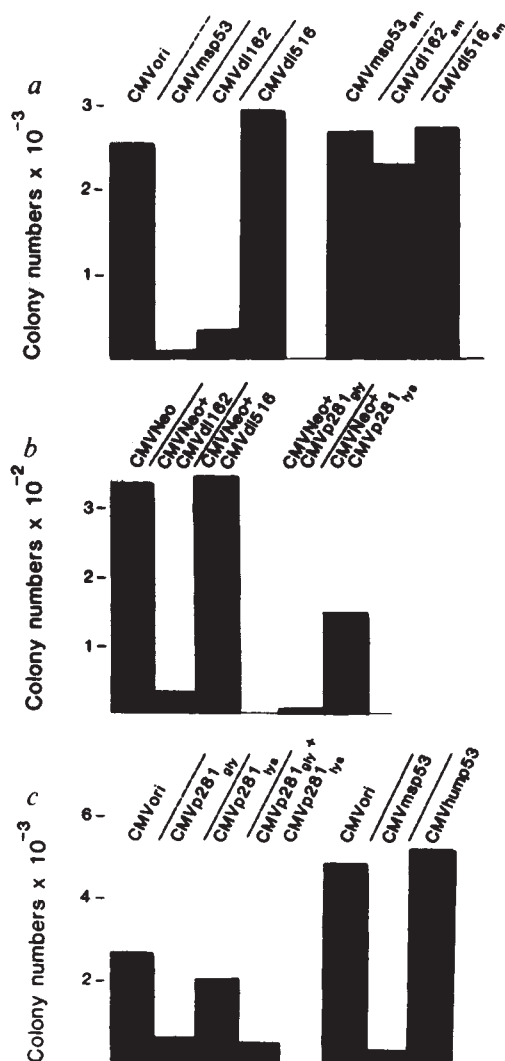
1. Rigby, P. W. J. & Lane, D. P. in *Advances in Viral Oncology* Vol. 3 (ed. Klein, G.) 31-57 (Raven, 1983).
2. Lane, D. P. & Crawford, L. V. *Nature* **278**, 261-263 (1979).
3. McCormick, P. & Harlow, E. *J. Virol.* **34**, 213-224 (1980).
4. Harlow, E., Pim, D. C. & Crawford, L. V. *J. Virol.* **37**, 564-573 (1981).
5. Eliyahu, D., Raz, A., Gruss, P., Givol, D. & Oren, M. *Nature* **312**, 646-649 (1984).
6. Jenkins, J. R., Rudge, K. & Currie, G. A. *Nature* **312**, 651-654 (1984).
7. Parada, L. F., Land, H., Weinberg, R. A., Wolf, D. & Rotter, V. *Nature* **312**, 649-651 (1984).
8. Smaie, S. T. & Tjian, R. *Molec. cell. Biol.* **6**, 4077-4087 (1986).
9. Mole, S. E., Gannon, J. V., Ford, M. J. & Lane, D. P. *Proc. R. Soc. B* (in the press).
10. Lane, D. P. & Gannon, J. V. in *Cancer Cells* Vol. 4 (eds Botchan, M., Grodzicker, T. & Sharp, P. A.) 387-393 (Cold Spring Harbor, New York, 1986).
11. Tanaka, S., Hu, S., Wang, T. & Korn, D. *J. biol. Chem.* **257**, 8386-8390 (1982).
12. Stillman, B. W. & Gluzman, Y. *Molec. cell. Biol.* **5**, 2051-2060 (1985).
13. Stillman, B. W., Dille, J. F. X., Prelich, G. & Guggenheimer, R. A. in *Cancer Cells* Vol. 4 (eds Botchan, M., Grodzicker, T. & Sharp, P. A.) 453-463 (Cold Spring Harbor, New York, 1986).
14. Simanis, V. & Lane, D. P. *Virology* **144**, 88-100 (1985).
15. Ottiger, H., Frei, P., Hassig, M. & Hubscher, U. *Nucleic Acids Res.* **15**, 4789-4807 (1987).
16. Lane, D. P., Gannon, J. & Winchester, G. *Adv. viral Oncol.* **2**, 23-39 (1982).
17. Reich, N. C. & Levine, A. J. *Nature* **308**, 199-201 (1984).
18. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861-869 (1981).
19. Benchimol, S., Pim, D. & Crawford, L. V. *EMBO J.* **1**, 1055-1062 (1982).
20. Yewdell, J. W., Gannon, J. V. & Lane, D. P. *J. Virol.* **59**, 444-452 (1986).
21. Rotter, V., Boss, M. A. & Baltimore, D. *J. Virol.* **38**, 336-346 (1981).
22. Dippold, W. G., Jay, G., DeLeo, A. B., Khoury, G. & Old, L. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1695-1699 (1981).

Mouse p53 inhibits SV40 origin-dependent DNA replication

Antony W. Braithwaite, Horst-Werner Sturzbecher, Christine Addison, Colin Palmer, Karran Rudge & John R. Jenkins

Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 0TL, UK

p53 is a cellular phosphoprotein that is present at elevated concentrations in cells transformed by different agents¹⁻⁵. p53 complementary DNA expression-constructs immortalize primary cells *in vitro*^{6,7} and co-operate with an activated *ras* oncogene in malignant transformation^{6,8,9}. Several reports have implicated p53 in mammalian cell cycle control and specifically with events occurring at the G₀-G₁ boundary^{10,11}. p53 forms specific complexes with simian virus 40 (SV40) large-T antigen^{1,2}, and such complexes are found associated with both replicating and mature SV40 DNA in lytically infected cells¹². In an accompanying paper Gannon and Lane¹³ report that in *in vitro* plate-binding assays, mouse p53 can displace polymerase α from complex with T-antigen. We have examined the *in vivo* consequences of expressing wild-type and mutant p53 proteins from other species in SV40-transformed



monkey cells. We report here that expression of mouse p53 results in a substantial and selective inhibition of SV40 origin-dependent DNA replication. In addition to any function in the G₀-G₁ transition, the data presented suggest that p53 may affect directly the initiation or maintenance of replicative DNA synthesis.

Replication experiments were carried out by transfecting SV40 origin-containing plasmids into the SV40-transformed monkey COS cell line¹⁵ which is permissive for SV40 replication. The plasmids are described in the legend to Fig. 1. Figure 1a shows that plasmids CMVmsp53 and CMVd1162 are restricted in their ability to replicate whereas CMVd1516 compares with the control CMVori. In five independent experiments, CMVmsp53 replicated on average only 4% compared with CMVori, and CMVd1162 remained slightly (3 to 4 times) higher than CMVmsp53. In another set of experiments, COS cells were transfected with wild-type and mutant p53 expression constructs containing an additional amber mutation at residue 609. The results of these transfections are also shown in Fig. 1a. The amber mutations give complete restoration of the replicative capacity of both CMVmsp53 and CMVd1162. CMVd1516^{am} remains essentially identical to CMVd1516. We conclude that expression of a full-length encoded protein product is required for restriction of DNA replication. p53 expression constructs were next co-transfected into COS cells together with the con-

Fig. 1 Mouse p53 inhibits SV40 origin-dependent DNA replication.

Methods. Monkey COS cells are seeded from confluent flasks into 60 mm tissue culture dishes in Dulbecco's modified Eagle's medium (DMEM; Flow Laboratories) supplemented with 10% fetal calf serum (FCS; Flow Laboratories) and L-glutamine. 2 h later cells are washed in DMEM minus FCS and transfected with 5 µg of plasmid DNA per dish using the DEAE/Dextran procedure²⁴. After a further 4 h, cells are treated with Hank's buffered saline containing 10% dimethylsulphoxide for 2 min. Cells are washed and then incubated for 72 h in DMEM + 10% FCS at 37 °C with 10% CO₂, washed with phosphate-buffered saline (PBS) and the plasmid DNA isolated using modified Hirt supernatant procedures³⁰. Plasmid DNA is digested with excess *DpnI* restriction endonuclease (New England Biolabs). Parent and replicated daughter DNA molecules are distinguishable by their differential sensitivity to the methylation sensitive enzyme *DpnI* (ref. 16). The restriction enzyme digestions are used to transform competent DH5 *E. coli*³¹. *DpnI* resistant DNA molecules representing the replicated fraction transform *E. coli* efficiently and are thus a measure of plasmid replication¹⁶. Transformation efficiency was about 10⁷-10⁸ bacterial colonies per µg input plasmid. Cells are plated on L-agar supplemented with 0.01 M MgCl₂ and 50 µg ml⁻¹ ampicillin or kanamycin, and incubated overnight at 37 °C. Colonies are counted manually. Plasmids: all p53 cDNA expression vectors used in the present study were constructed in the pBC12CMV vector described by Cullen¹⁴. This contains the 200 base pair *SphI* to *HindIII* restriction fragment from SV40 DNA which encompasses a functional viral DNA replication origin. CMVori lacks an expressed p53 cDNA gene; CMVmsp53 encodes wild-type mouse p53; CMVd1162 and CMVd1516 contain internal deletions of mouse p53 (see ref. 7); CMVmsp53^{am}, CMVd1162^{am} and CMVd1516^{am} each contain an additional amber mutation at residue 609 (ref. 20) which causes translation to terminate at the amber codon, giving rise to a truncated N-terminal peptide (J.R.J. and C.P., manuscript in preparation); CMVNeo encodes the Tn5 gene product that confers resistance to the antibiotic G418 in animal cells, and to the neomycin-kanamycin family in *E. coli*²¹; CMVp281^{lys} and CMVp281^{lys} express mouse p53 proteins in which the normal amino acid at position 281 has been converted respectively from glutamic acid to glycine and lysine (C.A. and J.R.J., in preparation). In several experiments, parallel transfections were carried out in which cells were fixed and examined for p53 expression by indirect immunofluorescence. Using separately the murine specific monoclonal antibodies pAB242 and pAB248 (ref. 18) and the rodent specific monoclonal antibody pAB200-47 (ref. 19), CMVd1516, CMVmsp53 and CMVd1162 all showed ~30-40% p53 positive cells.

struct CMVNeo, which encodes kanamycin resistance. The results (Fig. 1b) show that CMVd1162 suppresses replication of the cotransfected CMVNeo, whereas CMVd1516 has no effect. Similar results were obtained with two point mutants of p53 (see below, and Fig. 1b). We conclude that the restriction of replicative capacity shown by CMVd1162 operates in *trans* and, together with the results of the amber mutant experiments, argues that it is directly due to the encoded mouse p53 protein.

To eliminate the involvement of some non-specific toxic effect, COS cells were transfected with CMVori, CMVmsp53 or CMVd1516 and these cell populations examined by flow cytometry²². The results (Fig. 2) show that there is no detectable alteration in the distribution of DNA contents. Similar results were obtained by pulse-labelling replicating DNA with ³H-thymidine (data not shown). We conclude that there is no significant inhibition of cell cycle progression or of cellular DNA synthesis by the mouse wild-type cDNA gene product, and that restriction of replicative synthesis is therefore SV40 origin-specific. Western blotting shows that there is no relationship between steady state levels of the expressed p53 proteins and inhibition (data not shown).

The mouse p53 expression-constructs described here fall into two classes: those encoding a T antigen-binding protein inhibit plasmid replication, but those encoding a non-T antigen-binding

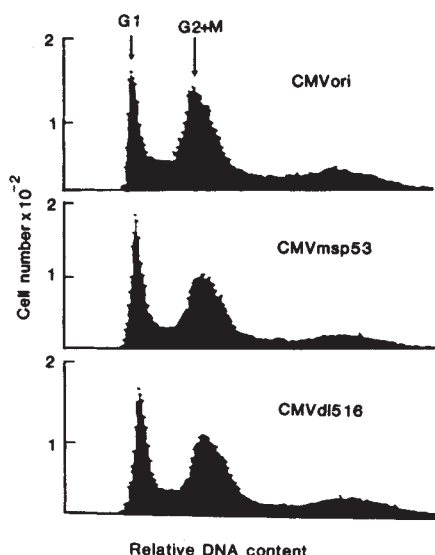


Fig. 2 Mouse wild-type and mutant p53s do not inhibit proliferation of transfected monkey COS cells. Cells were transfected as described for Fig. 1. 72 h post-transfection, cells were trypsinized and fixed in cold (-20°C) 70% ethanol and stored at -20°C . Just before analysis of DNA content by flow cytometry, fixed cells were briefly pelleted then resuspended in PBS and incubated with $100\text{ }\mu\text{g ml}^{-1}$ ribonuclease A for 30 min at 37°C , and the DNA stained with propidium iodide²². Under the conditions of the experiments described here in which 30–40% of cells were p53 positive by immunofluorescence (see legend to Fig. 1), previous work with a similar proportion of antigen positive cells showed substantial alterations in cell cycle progression by flow cytometry²³. The cell cycle phases G1, G2 and M (mitosis) are indicated.

protein^{24,25} do not. Two point mutants of mouse p53 were tested for their effects on plasmid replication: CMVp281_{gly} p53 which binds T-antigen, and CMVp281_{lys} p53, which does not. The results are shown in Fig. 1c. CMVp281_{gly} inhibits plasmid replication but CMVp281_{lys} does not, and in cotransfections inhibition is dominant (Fig. 1c). To determine if inhibition of replication is a general feature of T antigen binding-p53 protein overexpression, we made and transfected an expression construct encoding a full-length human p53 protein. Human p53 also binds T-antigen in these assays. Fig. 1c shows the result. Although the human protein is efficiently expressed, is detectable by indirect immunofluorescence in an equivalent percentage of the transfected cell population, and accumulates at high levels in such cells, there is no detectable inhibition whatsoever of SV40 origin-dependent replicative synthesis. It appears therefore that suppression of such replicative activity is a characteristic of the murine p53 protein and is not a general feature of the T/p53 interaction *per se*.

SV40 DNA replication requires the virally encoded T antigen²⁶ as well as a range of host cell proteins²⁹ and T-antigen/polymerase α complexes appear to be required for SV40 replication *in vitro*²⁷. p53 also forms complexes with large T antigen^{1,2} and in an accompanying paper¹³ it is demonstrated that mouse p53 can displace polymerase α from T-antigen in *in vitro* plate-binding assays. Taken in isolation, the mouse p53 data we present here, along with the observations of Gannon and Lane¹³, suggest that p53 might function as a regulatory factor controlling, for example, the recruitment of SV40 chromosomes for cycles of replicative synthesis. The human p53 result (Fig. 1c) suggests otherwise, however. It seems more likely that inhibition of replication by mouse p53, and perhaps DNA polymerase α displacement, also results from some intrinsic species difference between the way mouse and human (and presumably monkey) p53 interact with T antigen. For example, in SV40 permissive cells it could be DNA polymerase α that displaces p53 from a complex with T-antigen, and that p53 participates in some

intermediate step between T-antigen binding and DNA polymerase α -mediated synthesis. It is significant that mouse cells are non-permissive for SV40 replication whereas human cells are semipermissive²⁸, and SV40 replicates poorly in HeLa cells that lack a normal p53 protein. HeLa cell extracts do support SV40 DNA replication *in vitro*^{16,17}, yet not as efficiently as other human cell extracts¹⁷. We suggest that differences between mouse and human p53 are able to play a part in defining viral host range. A number of reports have implicated p53 in early events following the stimulation of quiescent cells, and specifically in the transit from G₀ to G₁^{10,11}. We suggest that p53 could also be directly involved in the initiation or maintenance of replicative DNA synthesis.

We thank Debbie Lynge-Hansen for tissue culture and immunofluorescence analysis, Ed Harlow for a human cDNA gene, Pat Imrie for flow cytometry, and Jean Marr for typing the manuscript. A. Braithwaite was supported by the Howard Florey Fellowship from Australia.

Received 13 July; accepted 24 August 1987.

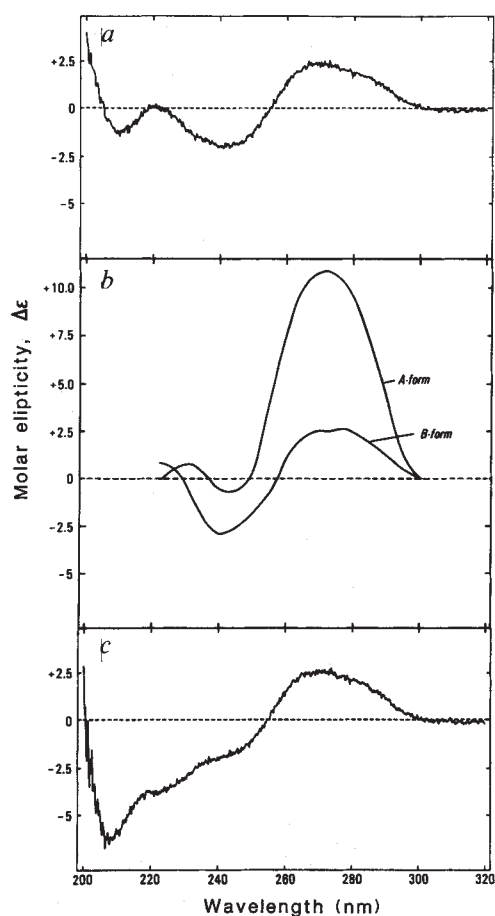
1. Lane, D. P. & Crawford, L. V. *Nature* **278**, 261–263 (1979).
2. Linzer, D. I. H. & Levine, A. J. *Cell* **17**, 43–52 (1979).
3. DeLeo, A. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 2420–2424 (1979).
4. Rotter, V., Boss, M. A. & Baltimore, D. *J. Virol.* **38**, 336–346 (1981).
5. Ruscetti, S. K. & Scolnick, E. M. *J. Virol.* **46**, 1022–1026 (1983).
6. Jenkins, J. R., Rudge, K. & Currie, G. A. *Nature* **312**, 651–654 (1984).
7. Jenkins, J. R., Rudge, K., Chumakov, P. & Currie, G. A. *Nature* **317**, 816–818 (1985).
8. Eliyahu, D., Raz, A., Gruss, P., Givol, D. & Oren, M. *Nature* **321**, 646–649 (1984).
9. Parada, L. F., Land, H., Weinberg, R. A., Wolf, D. & Rotter, V. *Nature* **312**, 649–651 (1984).
10. Mercer, W. E., Nelson, D., DeLeo, A. B., Old, L. J. & Baserga, R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6309–6312 (1982).
11. Reich, N. C. & Levine, A. J. *Nature* **308**, 199–201 (1984).
12. Tack, L. C., Wright, J. H. & Gurney, E. G. *J. Virol.* **58**, 635–646 (1986).
13. Gannon, J. V. & Lane, D. P. *Nature* **329**, 456–458 (1987).
14. Cullen, B. R. *Cell* **46**, 973–982 (1986).
15. Gluzman, Y. *Cell* **23**, 175–182 (1981).
16. Li, J. J. & Kelly, T. J. *Molec. cell. Biol.* **5**, 1238–1246.
17. Stillman, B. W. & Gluzman, Y. *Molec. cell. Biol.* **5**, 2051–2060.
18. Yewdell, J. W., Gannon, J. V. & Lane, D. P. *J. Virol.* **59**, 444–452 (1986).
19. Dippold, W. G., Jay, G., DeLeo, A. B., Khoury, G. & Old, L. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1695–1699 (1981).
20. Jenkins, J. R., Rudge, K., Redmond, S. & Wade-Evans, A. *Nucleic Acids Res.* **12**, 5609–5626 (1984).
21. Southern, P. J. & Berg, P. *J. molec. appl. Genet.* **1**, 327–341 (1982).
22. Traganos, F., Darzynkiewicz, Z., Sharpless, T. & Melamed, M. R. *J. Histochem. Cytochem.* **25**, 46–56 (1977).
23. Murray, J. D., Bellett, A. J. D., Braithwaite, A. W., Waldron, L. K. & Taylor, I. W. *J. cell. Physiol.* **111**, 89–97 (1982).
24. Sturzbecher, H.-W., Chumakov, P., Welch, W. J. & Jenkins, J. R. *Oncogene* **1**, 201–211 (1987).
25. Wade-Evans, A. thesis, Univ. of London (1986).
26. Campbell, J. L. A. *Rev. Biochem.* **55**, 733–771 (1986).
27. Smale, S. T. & Tjian, R. *Molec. cell. Biol.* **6**, 4077–4087 (1986).
28. Toozé, J. (ed.) *DNA Tumor Viruses* (Cold Spring Harbor, N.Y., 1980).
29. Murakami, Y., Wobbe, C. R., Dean, F. B., Weissbach, L. & Hurwitz, J. in *Cancer Cells* Vol. 4 (Cold Spring Harbor, N.Y., 1986).
30. Hirt, B. *J. molec. Biol.* **26**, 365–369 (1967).
31. Hanahan, D. *J. molec. Biol.* **166**, 557–580 (1983).

The 5S gene internal control region is B-form both free in solution and in a complex with TFIIA

J. M. Gottesfeld, J. Blanco & L. L. Tennant

Department of Molecular Biology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

Rhodes and Klug¹ have recently proposed that the internal control region of the *Xenopus* 5S RNA gene adopts an A-type DNA structure in solution. This suggestion was based on a Fourier analysis of both the spacing of DNase I cutting sites and on the distribution of G residues in the DNA sequence. Both revealed a ~5.6–5.7-base periodicity which the authors interpreted as a structural repeat every half helical turn of A-type DNA. This contention was strengthened by the finding that a 9-base-pair (bp) double-stranded deoxyoligonucleotide corresponding to residues +81 to +89 of the 5S gene exhibits an A' RNA-like crystal structure².



This region of DNA is of special interest as it forms the binding site for the 5S gene-specific transcription factor IIIA (TFIIIA)³. TFIIIA is a Zn^{2+} -binding protein which interacts with both the internal control region of the gene and the 5S transcript⁴⁻⁶. As base-paired regions of RNA are of the A type^{2,7}, it was reasonable to postulate that 5S DNA might also adopt this conformation¹. We report here that the circular dichroism (CD) spectrum of a synthetic 54-bp deoxyoligonucleotide corresponding to the TFIIIA binding site is similar to the CD spectrum of B-form DNA in solution. Further, DNA-TFIIIA complexes show an unaltered DNA CD component indicating no gross alteration in DNA structure on protein binding.

CD spectroscopy in the ultraviolet provides a very sensitive method of examining alterations in DNA conformation⁸. The magnitude of the positive CD band at 275 nm is highly dependent on helix winding angle and on base-pair twist⁹, two parameters which differ dramatically between the B and A types of DNA¹⁰. Correlations between X-ray structures of DNA and solution CD measurements are possible as CD spectra have been obtained from DNA films under the conditions of ionic strength and relative humidity used in the X-ray fibre diffraction studies of B- and A-form DNA¹¹. Figure 1a shows the CD spectrum of a synthetic 54-bp double-stranded deoxyoligonucleotide which corresponds in sequence to residues +44 to +97 of the *Xenopus laevis* 5S RNA gene. For comparison the CD spectra of B-form and A-form DNAs are shown in Fig. 1b (ref. 8). The oligonucleotide spectrum largely resembles that of B-form DNA in low ionic strength solutions or in wet films¹¹; however, there are some minor differences which are probably related to the fact that the 5S spectrum is of a unique sequence while the B-DNA spectrum is of calf thymus DNA and represents an average of all sequences in the genome. More importantly, major features are shared by both spectra: the magnitude of the major positive peak in both spectra: the magnitude of the major positive and negative peaks are

Fig. 1 Circular dichroism spectra of 5S DNA (a), B- and A-form reference spectra (b) and 5S DNA complexed with TFIIIA (c). CD spectra were obtained at a DNA concentration of $68 \mu\text{g ml}^{-1}$ (a) and $49 \mu\text{g ml}^{-1}$ (c) in 50 mM KCl, 10 mM Tris-Cl, pH 7.5, 5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 0.1 mM dithiothreitol using a 0.2-cm optical path length. Spectra were recorded with an AVIV 61DS spectropolarimeter scanning from 320 to 200 nm with data collection every 0.25 nm. The spectra shown (a and c) are the average of five independent measurements with base line corrected for the buffer alone. The molar ellipticity is plotted to take account for differences in DNA concentration. The reference spectra were taken from Sprecher *et al.*⁸.

Methods. Two 58-bp oligonucleotides were synthesized with an Applied Biosystems DNA synthesizer. These DNAs correspond in sequence to nucleotides +40 to +97 of the noncoding strand of the *X. laevis* 5S RNA gene¹⁹ and nucleotides +44 to +97 of the coding strand plus an additional four nucleotides of sequence 5'GATC3'. The oligonucleotides were deblocked and purified by gel electrophoresis. Equimolar amounts, estimated from absorbance at 260 nm, were mixed in a solution containing 0.1 M $NaHCO_3$, pH 8.0, and heated to 95 °C for 5 min and allowed to reassociate by slow cooling to 22 °C over 3-4 h. The duplex DNA was precipitated with ethanol, vacuum dried and resuspended in TFIIIA binding buffer (above). More than 97% of the DNA was double stranded as estimated by gel electrophoresis of an aliquot. TFIIIA was purified from 7S RNP particles from immature oocytes⁶, and separated from 5S RNA and other proteins by DEAE cellulose chromatography, ribonuclease A digestion and Bio Rex 70 chromatography¹³. The protein was homogeneous as judged by silver staining of SDS polyacrylamide gels. TFIIIA (120 μg) was added to 34 μg of the DNA in a final volume of 700 μl providing a 3.5 molar excess of TFIIIA over DNA. The sample was incubated at 22 °C for 40 min before recording the CD spectrum over the course of 45 min at 23 °C. SDS gel electrophoresis demonstrated that the TFIIIA remained intact during this incubation. Likewise, electrophoresis under nondenaturing conditions showed that the complex remained intact.

similar in magnitude in both the 5S and B-form DNA spectra (Fig. 1b and refs 8 and 11). This is clearly not the case for A-form DNA where the positive peak is approximately ten times more intense than the negative⁸.

To demonstrate further that the 5S DNA spectrum obtained in aqueous solution reflects a B-type DNA structure, we examined the effect of solvent composition on the spectrum. Figure 2 shows that the 5S DNA oligonucleotide exhibits an A-type spectrum in 80% ethanol; further, this spectral transition occurs between 70 and 80% ethanol in both 5S DNA (not shown) and in bulk genomic DNA^{8,11}. We previously measured the helical repeat of 5S DNA in solution using the topoisomer bandshift method¹² and found a value of ~ 10.6 bp per helical turn over the region of +28 to +97 of the 5S gene¹³. This value is typical of DNA in the B family of structures^{12,14}. Taken

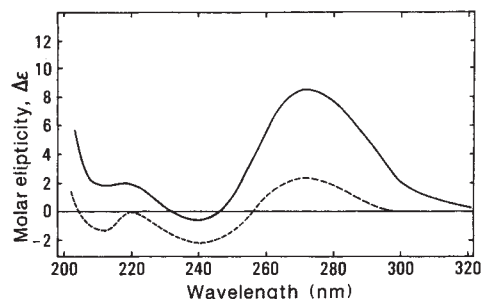


Fig. 2 B $\rightarrow$ A type transition in 5S DNA. CD spectra of the 54-bp double-stranded oligonucleotide were obtained in aqueous solution (10 mM Tris-Cl, pH 7.5, 0.1 mM EDTA, dashed line) and in 80% (v/v) ethanol (solid line). The DNA in aqueous solution was dialysed against steps of 25%, 50% and 80% (v/v) ethanol at 22 °C and insoluble material was removed by centrifugation. Molar ellipticity was calculated based on the absorbance of the supernatant at 260 nm.

together, these data and the present CD measurements support the view that the 5S internal control region is B-type in low ionic strength aqueous solution.

We have also determined the CD spectrum of the 54-bp oligonucleotide in a complex with TFIIA (Fig. 1c). In this experiment, a ratio of 3.5 molecules of TFIIA per 5S DNA molecule was used to ensure that all DNA molecules were complexed with protein. The spectrum of the complex above 240 nm is largely unchanged from that of the protein-free DNA. Below 240 nm the protein greatly influences the spectrum. These data suggest that no major structural change such as B→A transition occurs in the DNA upon TFIIA binding.

These data are thus in agreement with the results of nuclease digestions on the TFIIA-DNA complex where it has been found that the periodicity of DNase I cleavage within the TFIIA binding site is ~10.0–10.4 bases^{13,15}. If this spacing reflects the helical repeat of the DNA in the complex, then this supports the view that TFIIA binds DNA in a B-type structure. If no change in helical repeat occurs on TFIIA binding then previous measurements of change in linking number of TFIIA binding¹⁶ must indicate bending or wrapping of the DNA about the protein. Thus, our present results indicate that TFIIA can

recognize and bind to both B-type DNA and to A-type RNA¹⁷.

We thank Drs Peter Wright and David Stout for discussions and Peter Wright for the use of the CD spectropolarimeter and the NIH and the American Cancer Society for financial support.

Received 15 May; accepted 11 August 1987.

1. Rhodes, D. & Klug, A. *Cell* **46**, 123–132 (1986).
2. McCall, M., Brown, T., Hunter, W. N. & Kennard, O. *Nature* **322**, 661–664 (1986).
3. Engelke, D. R., Ng, S. Y., Shastry, B. S. & Roeder, R. G. *Cell* **19**, 717–728 (1980).
4. Miller, J., MacLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
5. Diakun, G. P., Fairall, L. & Klug, A. *Nature* **324**, 698–699 (1986).
6. Picard, B. & Wegnez, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 241–245 (1979).
7. Arnott, S., Hukins, D. W. L., Dover, S. D., Fuller, W. & Hodgson, A. R. *J. molec. Biol.* **81**, 107–122 (1973).
8. Sprecher, C. A., Baase, W. A. & Johnson, W. C. *Biopolymers* **18**, 1009–1019 (1979).
9. Johnson, B. B., Dahl, K. S., Tinoco, I., Ivanov, V. I. & Zhurkin, V. B. *Biochemistry* **20**, 73–78 (1981).
10. Arnott, S. & Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **47**, 1504–1509 (1972).
11. Tunis-Schneider, M. J. B. & Maestre, M. F. *J. molec. Biol.* **52**, 521–541 (1970).
12. Peck, L. J. & Wang, J. C. *Nature* **292**, 375–378 (1981).
13. Gottesfeld, J. M. *Molec. cell Biol.* **7**, 1612–1622 (1987).
14. Rhodes, D. & Klug, A. *Nature* **286**, 573–578 (1980).
15. Rhodes, D. *EMBO J.* **4**, 3473–3482.
16. Reynolds, W. F. & Gottesfeld, J. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1862–1866 (1983).
17. Christiansen, J., Brown, R. S., Sproat, B. S. & Garrett, R. A. *EMBO J.* **6**, 453–460 (1987).
18. Saenger, W., Hunter, W. N. & Kennard, O. *Nature* **324**, 385–388 (1986).
19. Peterson, R. C., Doering, J. L. & Brown, D. D. *Cell* **20**, 131–141 (1980).

Spatial relationship of the Fis binding sites for Hin recombinational enhancer activity

Reid C. Johnson*, Anna C. Glasgow†
& Melvin I. Simon†

* Department of Biological Chemistry, University of California at Los Angeles School of Medicine, Los Angeles, California 90024, USA

† Department of Biology 147-75, California Institute of Technology, Pasadena, California 91125, USA

Site-specific recombination reactions involve the joining or re-arrangement of discrete DNA segments in a highly precise manner. A site-specific DNA inversion regulates the expression of flagellin genes in *Salmonella* by switching the orientation of a promoter^{1,2}. Analysis of the reaction has shown that, in addition to DNA sequences at the two boundaries of the 1-kilobase invertible segment where strand exchange occurs, another *cis* acting sequence is required for efficient inversion³. This 60-base-pair enhancer-like sequence can function at many different locations and in either orientation in a plasmid substrate. It includes two binding sites for a host protein called Factor II or Fis (refs 4 and 5). Here we have investigated the importance of the spatial relationship between the two Fis binding sites for enhancer activity and have found that the correct helical positioning of the binding sites on the DNA is critical. However, this result could not be accounted for by effects on Fis binding. We propose a model for enhancer function in which the enhancer region acts to align the recombination sites into a specific conformation required for productive synapsis.

The inversion reaction can be performed in a purified *in vitro* system derived from *Escherichia coli*⁴. The Fis protein and two other proteins, Hin and HU, are required for maximum rates of inversion. The Hin protein, which is encoded by sequences within the invertible segment (see Fig. 1a), interacts specifically with the DNA sequences at the two 26-base-pair (bp) recombination sites (*hixL* and *hixR*). Hin is a member of a family of prokaryotic site-specific recombinases which promote DNA inversion, including the Gin (refs 6 and 7) and Cin (ref. 8) systems which control tail fibre expression in bacteriophage Mu and P1, respectively, and the Pin system⁹ whose function is unknown. The HU protein interacts nonspecifically with DNA and shows the greatest stimulatory effect on the reaction when

the enhancer is located within 350 bp of the closest recombination site⁴.

The recombinational enhancer sequence is normally located 103 bp from *hixL* (Fig. 1a), although it can function on either side of the recombination sites and as far as 4 kilobases (kb) from the closest recombination site³. Recombination rates of plasmid substrates lacking the enhancer are <0.5% of those found for plasmids containing enhancers. The nucleotide sequences responsible for enhancer function have been previously mapped by deletion and substitution analysis to extend from nucleotide +103 to +163 (as measured from the centre of *hixL*, see Fig. 1). Within this 60-bp sequence, the enhancer is divided into two domains referred to as the proximal and distal domains with respect to *hixL*. Footprinting and chemical interference experiments have demonstrated that the Fis protein binds to these two domains (Fig. 1b)²⁴. Fis binds with ~fivefold higher affinity to the *hixL* distal domain than to the *hixL* proximal domain.

To investigate the importance of the relative position of the two Fis binding sites for enhancer function, we constructed mutant enhancers containing increasing numbers of base pairs between the two domains. The insertions are all located at the *ClaI* restriction site between the proximal and distal domains and range in size from 2 to 43 bp (Fig. 1c). The mutant substrate DNAs were assayed in the reconstituted *in vitro* system⁴. Figure 2 shows the data for reaction times of 3 and 30 min. Under the conditions used in these experiments, the wild-type substrate pMS551 generates 50% of the final yield of recombinant molecules after ~2 min and the maximum yield of recombinants after ~15 min (data not shown). The longer incubation times were used for the data presented in Fig. 2 to enable the lower reaction rates found in the mutants to be measured.

Increasing the spacing between the enhancer domains by two base pairs (pMS553) results in an 80% decrease in the number of recombinants measured after a 3-min incubation period as compared with the wild-type substrate. After 30 min, slightly over half the wild-type number of recombinants have accumulated with pMS553. An insertion of 5 bp, or half a helical turn of DNA, abolishes enhancer activity as pMS680 recombines at a rate that is slightly lower than a plasmid in which the proximal domain of the enhancer is deleted (pMS609). However, an insertion of 10 bp (pMS552), or approximately one complete turn of the DNA helix, largely restores enhancer function (about 70% of the wild-type rate after 3 min). Plasmid pRJ719, which also contains a 10-bp insertion but with a different sequence from pMS552 (Fig. 1c), recombines at about 45% of the wild-

Fig. 1 *a*, Structure of the invertible DNA segment controlling *H2* expression in *Salmonella*. The *hin* gene, the left and right recombination sites (*hixL* and *hixR*), and the recombinational enhancer are shown. P, approximate position of the promoter which transcribes *H2* when oriented in the ON orientation as shown here. *b*, Nucleotide sequence around the *Hin* recombinational enhancer³. Bar, boundaries of the sequences from both strands that are protected from DNase I cleavage by Fis in footprinting experiments; *, purine which when methylated by dimethyl sulphate prevents binding by Fis²⁴. Shown above the sequence is a possible recognition sequence for Fis binding. *c*, Plasmids containing mutant enhancers: pMS551 contains the wild-type enhancer region³; the other plasmids are derivatives of pMS551 and contain the indicated number of base pairs inserted in the *ClaI* site at position 122. The sequence of the inserted DNA is shown in brackets.

Methods. Plasmid pMS551 has *hixL* and the enhancer region (positions -99 to +207) between the *EcoRI* and *HindIII* sites and a synthetically-derived 26-bp *hixL* site cloned into the *SalI* site of pBR322 (ref. 3). The construction of pMS553 and pMS552 has also been described³. Plasmid pMS680 was made by digesting pMS552 with *XhoI*, which cuts within the inserted DNA, and digesting with nuclease *S_I* (10 min at 30 °C with 1 unit *S_I* nuclease (Pharmacia) per μ g DNA) followed by blunt end ligation; pRJ719 was made by digesting pMS551 with *ClaI* and ligating in the presence of a 500-fold molar excess of the oligonucleotide pCGTTCAGAA. Plasmid pMS673 was constructed as follows: pMS551 was cleaved with *EcoRI*, the protruding ends filled in with T4 DNA polymerase and dNTPs, and religated to destroy the *EcoRI* site. This plasmid (pMS672) was then digested with *ClaI*, filled in with T4 DNA polymerase and dNTPs, and the *EcoRI*-site containing linker CCGGAATTCGGG was inserted to give pMS673. Plasmid pMS684 was derived from pMS673 by digestion with *EcoRI*, filling in with T4 DNA polymerase and dNTPs, and blunt end ligation; pMS696, pMS695 and pMS698 were constructed by ligating *XhoI*-cleaved pMS552 in the presence of a 500-fold molar excess of the oligonucleotide pTCGATCGA/TCGA giving rise to plasmids containing between one and three inserts of the oligonucleotide, respectively; pMS660 and pMS662 were made by ligating a 24-bp and 30-bp, respectively, *TaqI* fragment from pRZ102 (colE1::Tn5, ref. 19) into *ClaI* cleaved pMS551; pMS681 was constructed by ligating the *NcoI*-site-containing oligonucleotide AATCCATGG into *EcoRI* digested pMS673; and pMS683 was made by digesting pMS681 with *NcoI* and filling in the protruding ends with T4 DNA polymerase and dNTPs followed by blunt end ligation. Sequence analysis indicated that only three of the four bases of each end were filled in, with one base deleted, giving rise to the sequence shown for pMS683. The nucleotide sequence of the entire enhancer region for each of these plasmids was confirmed by direct plasmid dideoxy sequencing²⁰ using an oligonucleotide primer from +71 to +90.

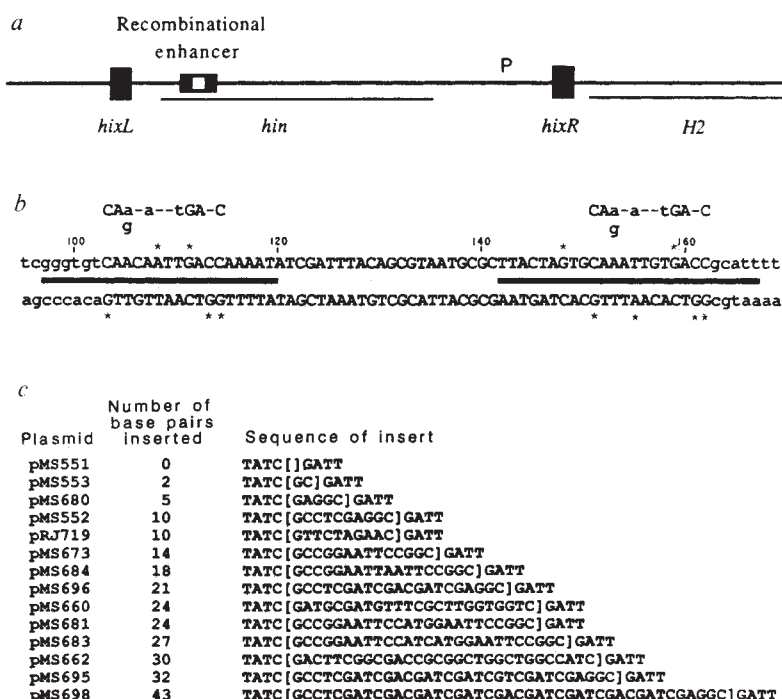
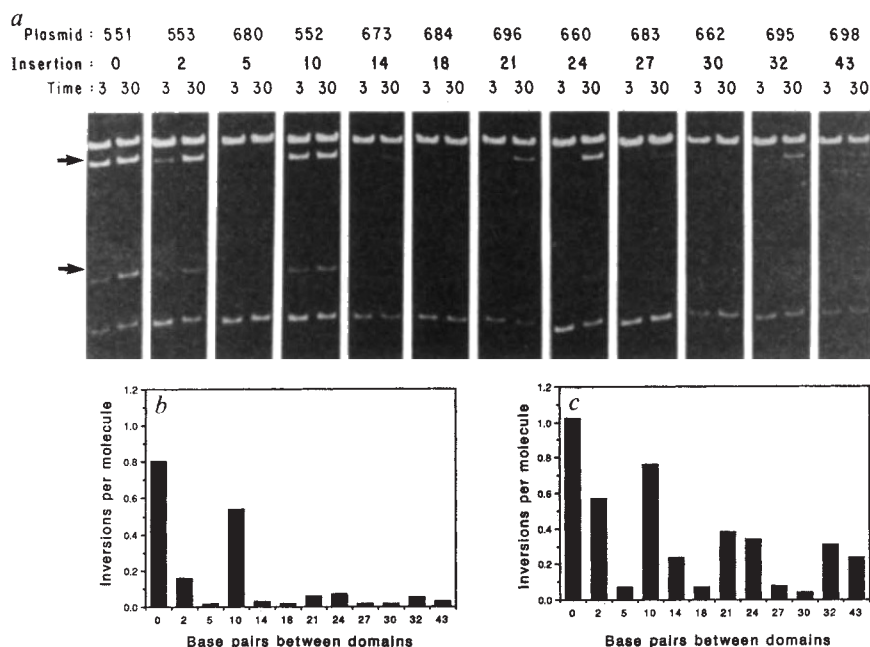


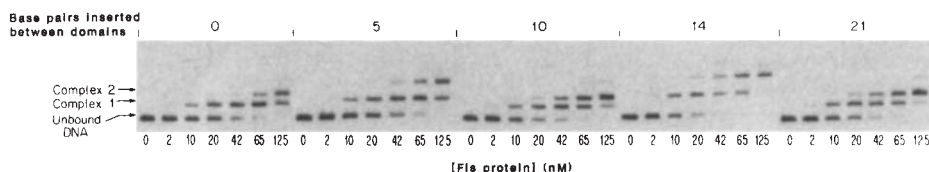
Fig. 2 *a*, Inversion assays *in vitro* of enhancer spacer mutants. Plasmid DNA was incubated at 37 °C with purified *Hin*, *Fis* and *HU* for 3 and 30 min as described⁴, then the mixture was digested with *PstI* and *HindIII* and analysed by agarose gel electrophoresis. Arrows, inversion products. The substrate plasmid, number of base pairs inserted, and incubation times are given above the gels. *b*, *c*, Bar graphs depicting the number of inversions per DNA substrate molecule for each plasmid. *b*, 3-min and *c*, 30-min assays. The proportion of molecules in the inverted orientation was determined by scanning photographic negatives with a laser densitometer and the number of inversions per DNA molecule was calculated²¹. Not shown are the results of pMS681 (+24), which gave essentially identical values as pMS660 (+24), and pMS609, which is deleted for the proximal domain of the enhancer³ and which gave <0.02 and 0.08 inversions per molecule after 3 and 30 min, respectively.



type rate and accumulates about 90% of the wild-type number of recombinants after 30 min. Insertions of 14 and 18 bp progressively decrease enhancer function whereas a 21-bp insertion, or two complete turns of the helix, results in significantly higher rates. This helical periodicity for enhancer function is maintained, albeit with very low reaction rates, for three (32 bp)

and probably four (43 bp) turns of the DNA helix. The recombination assays clearly indicate that the helical orientation of the two *Fis* binding sites in the enhancer is critical for function. Moreover, an increase of spacing of one complete turn has only a small effect on enhancer function, but two or more turns are poorly tolerated.

Fig. 3 Polyacrylamide gel electrophoresis of Fis/enhancer DNA complexes. A 276-bp *Hind*III-*Eco*RV restriction fragment from pMS551 containing the wild-type enhancer sequence and the corresponding *Hind*III-*Eco*RV restriction fragments from pMS680 (+5), pMS552 (+10), pMS673 (+14) and pMS696 (+21) were isolated after end-labelling at the *Hind*III site using the Klenow fragment of DNA polymerase and [α - 32 P]dATP. Binding and electrophoresis of the Fis/enhancer DNA complexes were as described²⁴. The binding reactions were performed in 80 mM NaCl, 25 mM Tris-HCl, pH 7.9, 1.2 mM EDTA, 2 mM dithiothreitol, 0.5 mg ml⁻¹ bovine serum albumin, 0.2 mg ml⁻¹ polycytidylic acid and 15 μ g ml⁻¹ sonicated calf thymus DNA. Electrophoresis times and conditions were identical for each titration.



Studies of repressor and activator interactions in several systems have shown the importance of correct helical orientation of protein binding sites¹⁰⁻¹². For example, λ repressor binding to variably spaced operator sites demonstrated that correct positioning on the helix was necessary to allow for cooperative interactions to occur between repressor proteins¹¹. In an analogous fashion, the helical dependence seen for Fis activity in recombination could reflect cooperative interactions between Fis proteins bound at the two domains of the enhancer. However, measurements of Fis binding show no evidence of cooperative interactions. Binding of Fis to the enhancer domains of the wild-type and selected mutant enhancers was measured using gel retardation assays^{22,23} (Fig. 3). Titration of Fis protein with a constant concentration of wild-type enhancer-containing restriction fragment revealed two bound species (complex 1 and complex 2 in Fig. 3). Complex 1 appears at low protein concentrations and consists of Fis bound predominantly at the distal domain; complex 2 contains Fis bound to both the proximal and distal domains²⁴. The insertion mutants show little differences in the formation of these complexes as a function of protein concentration. Additional experiments, where protection from DNase I cleavage was measured as a function of added Fis protein, resulted in no differences in affinity or footprinting patterns that correlated with the helical disposition of the Fis binding domains (data not shown). Although all these experiments are on linear DNA, the results suggest that the requirement for correct helical positioning of the enhancer domains on the DNA for Fis activity is not the result of cooperative interactions between Fis proteins which facilitate binding. This conclusion is further supported by the finding that DNA fragments containing isolated proximal or distal enhancer domains bind Fis with the same relative affinity as the wild-type fragment containing both domains²⁴. However, we cannot rule out the possibility that the helical dependence on recombination is a function of cooperative binding effects that only occur on superhelical DNA.

We have observed that the relative mobilities of the complexes in the gel retardation assay are different depending on the helical position of the Fis binding sites. Complexes formed with pMS680 (+5) and pMS673 (+14) enhancer DNA fragments show a greater electrophoretic shift than do complexes formed with pMS551 (wild-type), pMS552 (+10), and pMS696 (+21) DNA fragments (Fig. 3). The different mobilities may reflect unique structural features of the enhancer DNA¹³. DNA fragments containing mutant enhancers without bound Fis also migrate aberrantly on polyacrylamide gels (data not shown), suggesting deformations associated with the DNA helix. Binding by Fis protein may generate further structural changes in the DNA. Structural perturbations such as bends or kinks in the DNA would lead to altered relative mobilities depending on whether they are oriented towards, or away from, each other¹⁴. Thus, similar electrophoretic shifts are obtained with complexes generated with enhancer fragments that contain integral numbers of DNA helical turns between proximal and distal domains, whereas different shifts are obtained with fragments containing

insertions of a non-integral number of helical turns (Fig. 3).

Figure 4 shows a model for the function of the enhancer in site-specific DNA inversion. *Hin* and *Fis* bind independently to their respective recognition sites on the DNA. Protein-mediated interactions between *Hin* and *Fis* then occur which result in looping out of the intervening DNA between each recombination site and the enhancer. In this manner, the *Hin* molecules and recombination sites are assembled into the proper synaptic complex to give inversion of DNA after strand exchange.

A number of features of this model are consistent with the properties of the *Hin* enhancer. The enhancer/recombination site complex could assemble without regard to the precise location or orientation of the enhancer relative to the recombination sites. The only consideration in this respect is that there be sufficient distance between the protein binding sites for the DNA to be able to loop around to achieve the required protein interactions. Consistent with this, we have previously shown that enhancer cannot function when it is very close (within 30 bp) to one of the recombination sites³. We have also found that the major enhancing effect of the HU protein on the reaction occurs when the enhancer is within ~200 bp of the nearest recombination site⁴. The HU protein may be promoting or stabilizing the required bending of the DNA, and thus its stimulatory effect is greatest when the enhancer and recombination sites are close to each other.

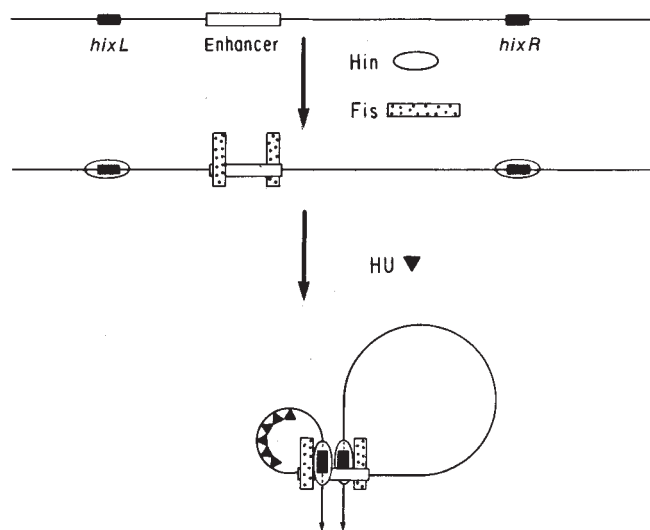


Fig. 4 Model for enhancer function in the formation of the synaptic complex. Subsequent to binding of *Hin* and *Fis* to their recognition sequences on the DNA, a tripartite nucleoprotein complex is assembled at the enhancer. The assembly of this complex could occur by independent association of each recombination site with an enhancer domain or, alternatively, *Hin* proteins could first pair the two recombination sites, followed by interaction with the enhancer.

In this 'looping-synapsis' model, the critical function of the enhancer is to position the recombination sites in exactly the correct configuration to complete synapsis. That enhancer function is dependent on the helical position of the Fis binding sites provides strong support for this idea. Separation of the enhancer domains by an additional half-turn abolishes enhancer activity, whereas function is regained when one full turn of the helix is inserted. Interactions which facilitate binding of Fis proteins at the two domains are probably not responsible for the observed helical dependence for enhancer activity because binding affinities measured on DNA fragments containing the altered spacing mutants are similar. Recent data suggest that bound Fis proteins are located in direct repeat orientation 48 bp from each other or ~ 4.6 turns of the helix²⁴. Thus, in Fig. 4 we have shown Fis proteins binding to opposite sides of the helix. Further studies, including analysis of the structural properties of the enhancer DNA, will be required before the precise spatial relationship of the two Fis binding sites is known.

The looping-synapsis model accommodates many of the features that have been found for the *Hin* enhancer and provides a useful guide in thinking about enhancers in site-specific recombination. An important prediction of the model is that specific protein-protein interactions occur between the *Hin* recombinase and Fis. We are currently probing for such interactions. It is also worth noting that the related site-specific recombinase, resolvase, requires two additional binding sites which are adjacent to the site of recombination¹⁵. The Fis binding sites in the *Hin* inversion system may have a function similar to those of the additional resolvase binding sites in aligning the protein and DNA sequences at the site of recombination^{16,17}. It seems

likely that the assembly of complex nucleoprotein structures involving multiple protein binding sites on the DNA will be increasingly recognized as a feature of protein/nucleic acid reactions in which distantly spaced regulator or enhancer sequences are involved¹⁸.

We thank Diana Pfeffer and Carol Lee for help with the DNA sequence analysis and Mike Bruist for helpful suggestions. This work was supported in part by a grant from the Weingart Foundation to R.C.J. and from the NSF to M.I.S. A.C.G. is a fellow of the American Cancer Society.

Received 24 July; accepted 4 August 1987.

1. Zieg, J., Silverman, M. & Hilman, M. *Science* **196**, 170-172 (1977).
2. Silverman, M. & Simon, M. *Cell* **19**, 845-854 (1980).
3. Johnson, R. C. & Simon, M. I. *Cell* **41**, 781-789 (1985).
4. Johnson, R. C., Bruist, M. F. & Simon, M. I. *Cell* **46**, 531-539 (1986).
5. Koch, C. & Kahmann, R. *J. biol. Chem.* **261**, 15673-15678 (1986).
6. Kamp, D., Kahmann, R., Zipser, D., Broker, T. R. & Chow, L. T. *Nature* **271**, 577-580 (1978).
7. van de Putte, P., Cramer, S. & Giphart, M. *Nature* **286**, 218-222 (1980).
8. Iida, S., Meyer, J., Kennedy, K. & Arber, W. *EMBO J.* **1**, 1445-1453 (1982).
9. Plasterk, R. H. A. & van de Putte, P. *EMBO J.* **4**, 237-242 (1985).
10. Dunn, T., Hahn, S., Ogden, S. & Schlieff, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5017-5020 (1984).
11. Hochschild, A. H. & Ptashne, M. *Cell* **44**, 681-687 (1986).
12. Takahashi, K. *et al. Nature* **319**, 121-126 (1985).
13. Wu, H. M. & Crothers, D. M. *Nature* **308**, 509-513 (1984).
14. Zinkel, S. S. & Crothers, D. M. *Nature* **328**, 178-181 (1987).
15. Grindley, N. D. F. *et al. Cell* **30**, 19-27 (1982).
16. Boocock, M. R., Brown, J. L. & Sherratt, D. J. in *DNA Replication and Recombination* (eds Kelley, T. J. & MacMacken, R.) 703-718 (Liss, New York, 1986).
17. Gellert, M. & Nash, H. *Nature* **325**, 401-404 (1987).
18. Echols, H. *Science* **233**, 1050-1056 (1986).
19. Jorgensen, R. A., Rothstein, S. J. & Reznikoff, W. S. *Molec. gen. Genet.* **177**, 65-72 (1979).
20. Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165-170 (1985).
21. Bruist, M. F. & Simon, M. I. *J. bact.* **159**, 71-79 (1984).
22. Garner, M. M. & Revzin, A. *Nucleic Acids Res.* **9**, 3047-3060 (1981).
23. Fried, M. G. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505-6525 (1981).
24. Bruist, M. F., Glasgow, A. C., Johnson, R. C. & Simon, M. I. *Genes Dev.* (in the press).

Can the product of the θ gene be a real globin?

J. B. Clegg

MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

A new member ($\theta 1$, or $\psi\alpha$) of the α -globin gene family has recently been identified in a number of species¹⁻⁵. In higher primates the $\theta 1$ gene has all the structural features apparently necessary for expression, and it appears to have long been under strong selective constraints which suggests that it could still be, or recently have been, a functional gene¹. No corresponding 'globin' has yet been identified, however. In some other species, galago⁵ and rabbit⁴ for example, the $\theta 1$ and $\psi\alpha$ genes have accumulated enough inactivating mutations for them to be considered genuine pseudogenes. Horses also have an α -like gene ($\psi\alpha$) (ref. 3), in a 3' position identical to the other species in relation to the functional α genes, and this also appears to have the elements required for a functional gene. The predicted amino-acid sequence, however, suggests that any 'globin' product is likely to be non-viable because it has a number of seriously deleterious amino-acid replacements. Some of these amino-acid changes are shared with the rabbit and primate sequences, indicating that they predate the mammalian radiation, and that if indeed any of these genes are still functional, they are unlikely to be making haemoglobin.

The sequence of the equine $\psi\alpha$ gene (Fig. 1) has normal initiation and termination signals, correct splicing sequences, the same poly(A)-addition signal (AGTAAA) as the rabbit and primate genes^{2,4}, and lacks any obvious inactivating mutations. As in the primate genes, the promoter region is displaced ~ 150 base pairs to 5' (bp) relative to its position in known functional α genes, and it lacks a canonical CCAAT sequence, the closest candidate being the ACAAT ~ 30 bp 5' to the ATA sequence at nucleotide 64.

Comparison (Fig. 1) of this sequence with that of the adjacent $\alpha 1$ gene shows that 71 (50%) of codons have had base substitutions, of which 53 (37%) are changes (replacements) that would produce an amino-acid substitution. Moreover, there are 22 codons in the horse $\alpha 1$ sequence that can change to terminators by a single base change; 11 have sustained changes in $\psi\alpha$, but none to a terminator codon, whereas 3-4 might have been expected on a random basis in the absence of any selective constraints. When expressed as a corrected⁶ percentage of the nucleotide sites that have had replacement (23%) or silent (44%) site changes, these figures are very similar to those seen in the orang-utan $\alpha 1$ - $\theta 1$ comparison² (Table 1). In contrast, since the divergence of horses and orang-utans, their α genes have sustained only 9% replacement site changes compared with 40% silent site changes.

In globin genes and pseudogenes of various species, replacement sites appear to undergo substitutions at the rate of approximately 0.1×10^{-9} per site per year, and silent and non-coding sites at 5-7 times this rate^{6,7}. Thus the 9% replacement site rate for the horse/orang α gene comparison equates roughly to their divergence at around the time of the mammalian radiation ~ 85 Myr ago. Applying similar considerations to the horse and

Table 1 Per cent corrected divergence of gene sequences

	Replacement sites	Silent sites
α genes		
Horse versus orang-utan	9	40
Rabbit versus human	12	29
Horse versus human	7	40
$\psi\alpha$ (θ) genes		
Horse versus orang-utan	12	23
Rabbit versus orang-utan	14	26
α versus $\psi\alpha$ (θ)		
Horse	23	44
Orang-utan	26	45

See ref. 6.

CCAGATGCGTGCTAAATTCATGTCATCTACTAACAATAACACTTTAAATATTCGTT
TCCATACAGCAAAACACCGAAAATGGCGGCTCGGGCTCCAGGGTGATTGGAAAGTTGGC 120
AGGCCGAGCAGACGCTCAGTCGAGCGCTCCCGGACCCCTCGGGACTGGACTCTGCGAGC
AGAGGCGCGGCGCGCGGAGTGGCGCGCGGAGGCGCGGACCCGCGCCAGGGGCG 240

AlaLeuAlaAlaAlaAspArgAlaThrVal
CAGCGCGCTCGCGGCGCGGCTCACGATGGCGGCTGGCGCGGACAGGCGACCGTGT
T T T C C A A C A C
Val Ser LysThrAsn

ArgAlaLeuTrpLysLysMetGlySerAsnValGlyValTyrAlaThrGluAlaLeuGlu
CGCGCGCTGGAAAGATGGGTAGCAACGTCGGCGCTACGCGACCGAGGCCCTGGAG 360
AAG CGCC GT G T CG C CT AG T GCG A A
Lys Ala Ser Val GlyHisAla Glu GlyAla

Ar
AGGTGCGGAGGCGCGCGCTCCGGCTCGCCCTCTTTGCTCCCAAGCCCTCCCGCC

gMetPheLeuGlyPheProSerThrThrT
CGGACGACCTCTCTCGCGTTCTCTCGCAGGATGTTCTGGGCTTCCCTCCACACGA 480
A A
Thr Lys

hrTyrPheLeuHisLeuAspLeuSerLeuGlySerThrGlnValLysAlaHisGlyGlnL
CTTACTCTCTCCACCTGGACCTGAGCTCGGCTCCACCCAGTCAAGCCACGCGCAGA
C T C T A G G A
Pro Phe His Ala Lys

ysValAlaAspAlaLeuThrLeuAlaValGluHisLeuGluAspLeuProArgAlaLeuS
AGTGGCGCGACGCGCTGACCTCGCGCTGAGCACCTGGAAGACCTGCCCGCGCCCTGT 600
G T GC C TG
Gly Gly Asp Gly

erAlaLeuArgHisArgHisValArgGluLeuArgValAspProAlaSerPheGln
CGGCTCTGCGCCACCGGCACCTTCGCGAGCTGCGAGTGGACCCGCGCAGCTTCCAGGTGA
AA A G T CA A A C T A A
Asn SerAspLeu AlaHisLys ValAsn Lys

GCCCGCGGGCGGGCGGGACCGCTGCTCCACGACGGGACGCGGGGACGACGCGGACC 720

LeuLeuGlyHisCysLeuLeuValThrProAl
CCGGGTGGGACGACCCCTCTCTCCGAGCTCTGGGCCACTGCCTGCTGGTGACCCCGC
T A T T TCC TTG
Ser Ser Leu

aArgHisPheProGlyAspPheSerProThrLeuHisAlaSerLeuValLysPheLeuSe
CCGGCACTTCCCGGGGACTTCAGCCTACCCCTGACGCGCTCGTGGTCAAGTTCTCTGAG 840
GTC C AAC T C G G C C A A T
Val Leu Asn Thr AlaVal Asp

rHisValIleSerAlaLeuAlaSerAspCysArg***
CCACGTGATCTCGGCGCTGGCCTCCGACTGTGCTAAACCGGGCTCGGTGGGTGCGGGG
AGT G A C T A A A A T
Ser SerThrVal Thr LysTyr

CCCCCGGCTGCCCTCCCGCGAGGTTCACGCTCGGAGTAAAGTCCCAAAAGCAAGCCC 960
GGTGGCGCTGAGCGGTCTCTCTCGGCTCGGCGGTGGGCGCTCTGCCCCCAGAGGAAGA
GAGGGTCTGGGAGACCTGAGCCACAGCGCTCACTGTACATCCGGGACCCCTGGGTGG
TCGAGGAGAGGTTTCAGTTCGCCCCACCCCTGGGCGCTGAGGCGAGTTTGGGAAGGCGG 1141

Fig. 1 Nucleotide sequence of the sense strand of horse $\psi\alpha$ gene with (above) the translated protein sequence. Below the main sequence are the nucleotide and amino-acid changes that characterize the coding regions of the horse $\alpha 1$ gene and its expressed globin. A 1.5-kb *Mst*III fragment of plasmid pGG2 (ref. 3), containing the entire $\psi\alpha$ gene, was digested with *Hae*III, and a library of the resulting fragments constructed in the sequencing vector pGV403. DNA from size-selected subclones was sequenced on both strands by the method of Maxam and Gilbert⁹ and aligned and overlapped with a partial sequence previously obtained³ by sequencing from known starting points on the restriction map of the $\psi\alpha$ gene region.

orang-utan α versus $\psi\alpha$ ($\theta 1$) comparison suggests a divergence time of ~250 Myr, assuming that $\psi\alpha$ and $\theta 1$ genes have been functional globin genes throughout this period. On the other hand, the number of silent site changes are roughly the same in each of these comparisons, though not reaching values generally considered to be saturated through multiple substitutions^{6,7}. Although the discrepancy between replacement and silent substitution rates might be explained by the inherent uncertainties in the calculations of evolutionary divergence times, another possibility could be that the assumption that the $\psi\alpha$ and $\theta 1$ are globin genes is incorrect. Perhaps for much of the time since divergence from α , the $\psi\alpha$ and $\theta 1$ genes were rapidly accumulating amino acid replacement changes as they

Table 2 Important functional residues in α globin that are replaced in putative $\psi\alpha$ or $\theta 1$ protein

Animal	Residue	Substitution	Predicted consequence to structure
HBO	B6(25)	Gly → Ala or Thr	Glycine B6 makes a short contact with Gly E8. Insertion of a side chain forces helices B and E apart
BO	B13(32)	Met → Thr	Met B13 is an essential haem contact
HBO	C3(38)	Thr → Ser or Ala	Thr C3 is at the $\alpha_1\beta_2$ contact and is essential for the T→R switch
H	C5(40)	Lys → Thr	Lys C5 is essential for the stability of the T-structure
H	F5(84)	Ser → Arg	Ser F5 is essential for the stability of the tertiary structure. There is no room for an Arg
H	F7(86)	Leu → Arg	Leu F7 is an essential haem contact
HO	G4(97)	Asn → Ser	Asn G4 is an essential haem contact
B	G6(99)	Lys → Pro	Proline would disrupt the G-helix
H	G16(109)	Leu → Pro	Proline would disrupt the G-helix
H	HC2(140)	Tyr → Cys	Tyr HC2 is essential for the stability of the tertiary structure

H, horse; B, olive baboon; O, orang-utan.

evolved to fulfil a new function. Their divergence could thus have occurred much more recently than the straightforward $\alpha 1$ - $\theta 1$ ($\psi\alpha$) replacement site comparison suggests.

The predicted amino-acid sequence of the horse $\psi\alpha$ protein suggests that this possibility may need further investigation. Of the 53 changes in the $\psi\alpha$ protein sequence compared with horse α -globin, at least eight are in key structural regions and would severely impair the stability of a globin subunit, haem binding, or the R→T structural transition that accompanies oxygenation and deoxygenation (Table 2). These considerations suggest that the horse $\psi\alpha$ protein, if it exists, is unlikely to function successfully as a haemoglobin subunit. As the primate $\theta 1$ protein sequences have some of these inactivating substitutions, as well as others, there must also be some doubt about their functional properties.

A search for transcripts in various tissues could help to clarify doubts about the transcriptional competence of the $\psi\alpha$ and $\theta 1$ genes⁸, but will not provide much insight into their function. It is worth noting that humans completely lacking the $\theta 1$ gene have been born live¹⁰, suggesting that if the $\theta 1/\psi\alpha$ genes have a function it is not one vital to fetal survival.

I thank Max Perutz for advice, and Linda Roberts for preparing the manuscript. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00284.

Received 8 July; accepted 28 August 1987.

- Marks, J., Shaw, J.-P. & Shen, C.-K. *Nature* **321**, 785-788 (1986).
- Shaw, J.-P., Marks, J. & Shen, C.-K. *Nature* **326**, 717-720 (1987).
- Clegg, J. B., Braend, M. & Goodbourn, S. E. Y. *Nucleic Acids Res.* **12**, 7847-7858 (1984).
- Cheng, J.-F., Raid, L. & Hardison, R. C. *J. biol. Chem.* **261**, 839-848 (1986).
- Sawada, I. & Schmid, C. W. *J. molec. Biol.* **192**, 693-709 (1986).
- Perler, F. et al. *Cell* **20**, 555-566 (1980).
- Elstratidis, A. et al. *Cell* **21**, 653-668 (1980).
- Proudfoot, N. *Nature* **321**, 730-731 (1986).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Fischel-Ghodsian, N., Beyer, E. C. & Higgs, D. R. *Nature* **329**, xxx-xxx (1987).

Plastic disposables in the laboratory

R. Conway and T. Chaplin

Plastic disposables have become essential equipment in most medical and industrial laboratories. The benefits they offer in cost, sterility and safety suggest that this trend is likely to continue.

THE large volume of samples generated by the increase in testing and analysis work in modern laboratories has created an increased demand for disposables. The glass 'laboratory durables' used for many years had a guaranteed life expectancy of ten years, and often lasted for fifty years or more. Today's plastic laboratory disposables, in contrast, are designed to be used once and then be thrown away.

By the time plastic Petri dishes and other early laboratory disposables began to appear on the market in the early nineteen fifties, the concept of disposability was already well accepted. It was the arrival of auto-analysis equipment designed specifically for use with plastic or glass disposable 'accessories', however, which started the current era of laboratory disposables.

The properties of early plastics were very limited. While some were flexible enough to be formed into a variety of shapes, they lacked heat resistance and clarity, and this restricted their potential applications. The development of more sophisticated tools and machinery for manufacturing plastics brought the production of plastics from a primitive chemical process to an industry of high precision engineering, with the capability to design and mould accessories to almost any specification. Advances in plastics technology have now created a new generation of disposables with better resilience, heat resistance, clarity and surface properties that compete with traditional disposables made of glass. The range of plastic disposables available today includes such basic items as Pasteur pipettes and test tubes, as well as precision pipettes and centrifuge tubes that can withstand up to 8,000 RCF.

Composition

The majority of plastic disposables are moulded or extruded from polystyrene, polypropylene and polyethylene (Table 1). Polypropylene differs from polystyrene and polyethylene because it is autoclavable and has the ability to withstand temperatures of up to 135 °C. Polypropylene and high density polyethylene can both withstand chemical sterilization with substances such as benzalkonium chloride, formalin and ethanol. Although polystyrene cannot stand up to extreme heat and strong chemicals, it has a high level of optical clarity.

Another new plastic, polycarbonate, is extremely resilient and has excellent

optical clarity. Polycarbonate is classified as an engineering thermoplastic, because of its toughness, exceptional clarity and high heat resistance. It is made in an interfacial process by a reaction between bisphenol A and carbonyl chloride. Characteristics of this plastic include a heat deflection temperature of 130 °C and 87–91 per cent light transmission, and its impact properties make it ideal for centrifuge tubes.

One of the latest materials to be developed is very low density polyethylene (VLDPE) which can be drawn into thin film and does not tear easily. This new class of linear polyethylene polymer has a density range of 0.890–0.915 g cm⁻³. VLDPE has greater flexibility than the other low density linear polyethylenes — such as the lower strength material EVA and plasticized PVC — but does not provide their tough-

ness and broad operating temperature ranges. VLDPE can be used to produce thin films as low as 0.2 mm in thickness, and is suitable for applications requiring a soft, flexible material for disposables such as gloves and health-care products.

Precision

Disposables for liquid measuring and delivery, such as pipettes, require a high degree of precision in manufacturing to ensure that they are made to exact dimensions and will dispense exact volumes. Accordingly, precision plastics are made from high-quality components with specific constant properties.

The production of plastic disposables is backed up by strict quality control procedures. During manufacturing, pipette tips are subject to eleven different quality control procedures, and are screened according to uniformity of orifice geometry, cleanliness of hydrophilic surfaces, appropriate length and colour, and zero ovality. They are also tested for excess plastic residue, known as flash.

Plastics are able to deliver the precision required and still retain their benefits of low cost and disposability. A typical example is the micropipettor, where precision engineering is used to design the non-disposable plunger unit, while

contact with liquids is restricted to the disposable pipette tip. Tips are usually made of polypropylene, which can be polished to a high gloss to avoid retention of liquid which might prevent a precise delivery.

The latest techniques in plastic manufacturing are producing disposables with specific surface properties. This type of disposable is very useful in tissue culture, where the surfaces of culture dishes must be specially treated to encourage the growth of a uniform monolayer of cells. The treatment of plastics has also produced

Table 1 Physical properties of plastic disposables

	Polystyrene	Polypropylene	High density polyethylene
Autoclavable	No	Yes	No
Brittle temperature	0 °C	0 °C	– 100 °C
Chemical sterilization*	No	Yes	Yes
Dry heat sterilization at 180 °C	No	No	No
Gas sterilization, ethylene oxide	Yes	Yes	Yes
Heat distortion temperature	70 °C	135 °C	121 °C
Specific gravity	1.040	0.903	0.944
Transparency	Clear	Translucent	Opaque

*Benzalkonium chloride, formalin, ethanol.

multi-well plates which allow antigens to bind to their surface, making possible new immunological techniques such as ELISA and RIA.

Limitations

The applications of plastic disposables are limited by their moderate resistance to heat and chemicals. Although plastic tubing, for example, can be used with a strong base such as concentrated sodium hydroxide or a strong acid such as concentrated hydrochloric acid, it is not recommended for use with chemicals such as benzaldehyde, ethylene dichloride, acetic acid, ethers, concentrated sulphuric acid, lacquer solvents and perchloric acid. The manufacturers of plastic disposables provide safety guidelines for the storage and use of chemicals in conjunction with plastic products to prevent accidents arising from misuse.

The future in new plastics technology lies in the development of disposables to meet the changing requirements of the modern laboratory. The cost of raw materials and the introduction of stronger polymers will be determining factors in the future of disposables. □

Robin Conway and Tony Chaplin are at Elkay Laboratory Products Ltd, Unit 2, Crockford Lane, Basingstoke, Hants RG24 0NA, UK.

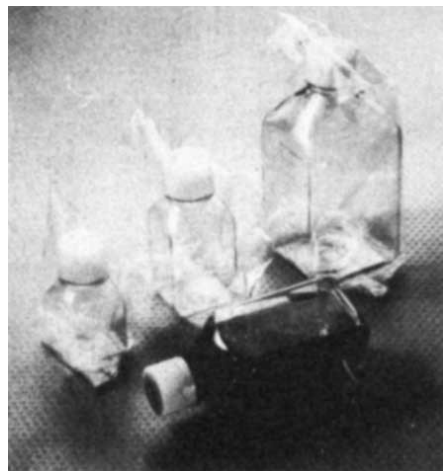
Disposables that make the grade

This week's feature on plastics and other disposables for laboratory use ends with a preview of several products to be exhibited next week at the American Society of Human Genetics meeting in San Diego, California.

FISHER now sells **microtome blades** in a hand-sized dispenser (*Reader Service No. 101*). The \$60 (US) dispenser comes with 60 stainless-steel blades measuring 0.498 in. wide by 3.000 in. long, that are compatible with AO blade-holders. The blades' 0.012 in. cutting edge is coated with PTFE, to allow tissue sections to slide off the blade without distortion and "chattering". The blade dispenser is designed for safety: the transparent top permits viewing of the remaining stock, and used blades are placed in the bottom compartment. The package can be disposed of as a unit after the blade supply is depleted.

The **1986/1987 catalogue** from Evergreen Scientific contains a full range of disposables for the laboratory or clinic (*Reader Service No. 102*). Available free from the company, the 64-page catalogue covers filters, pipettes and tips, tubes and caps, biohazard bags, vials, cuvettes, chromatography columns, and specimen containers and mailers. New in this catalogue are hospital supplies such as patient admission kits, pathology specimen containers, a sputum collection kit, and histology supplies. Evergreen offers free samples of many of its products.

Sterile cell culture media, serum and other biological solutions are just right for Nalge's **sterile disposable square bottles** (*Reader Service No. 103*). The polystyrene bottles have a deep polypropylene closure and a narrow mouth to facilitate aseptic techniques. The square design gives the bottles stability, and makes it easier to pack them into refrigerators and freezers, or line them up under a lab hood. The bottles are radiation-sterilized and pack-



Nalge's square container for sterile solutions.

aged in plastic, and have molded-in graduations for easy measuring. The bottles are available in 125, 250, 500 and 1,000 ml sizes, and are sold 48 and 24 to a case for \$48-72 (US).

The Super Chemisorb **masks** from Safelab Systems Ltd are now available in a more anatomically-moulded shape (*Reader Service No. 104*). Safelab says the new



Masks to guard against vapours and particles.

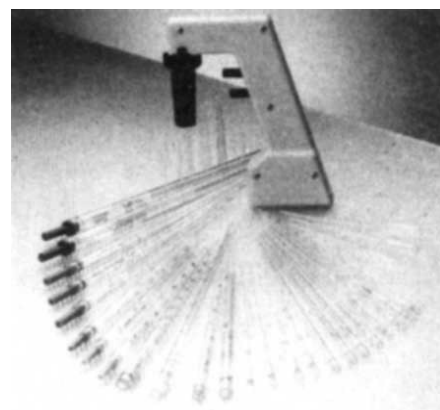
design fits better, to give greater protection from noxious fumes, mists, dust and smoke to the wearer. The Chemisorb masks have activated charcoal filter layers that absorb aromatic and aliphatic hydrocarbons, sulphur and nitrogen compounds, halogens, acids, alcohols, esters and ethers, and aldehydes and ketones. The masks also have a BS-6016 Type 2 particle filter to keep out particulate contaminants. The low breathing resistance of the masks permits normal speech. The masks come 10 to a box for £25 (UK) or 200 to a carton for £450 (UK).

For **cleaning up laboratory spills**, Matarah Industries has an on-the-spot solution, called Labdab (*Reader Service No. 105*). Labdabs are circular absorbent blotters backed with chemical-resistant polyester and a pinch-tab for handling. The Labdabs come in 3, 6, 9 and 12 in. diameter sizes. Matarah says they absorb up to 15 times their weight in liquid: the 3 in. Labdab absorbs up to 10 ml. The \$24-27 (US) Labdabs come 15 to a pack in a dispenser that can be mounted on a handy place near the lab bench.

World Precision Instruments offers **disposable ion-selective electrodes**, to avoid contaminating one solution with residue from a previous one (*Reader Service No. 106*). The Kwik-Tip disposable tips are 3 mm in diameter plastic tubes covered at one end with a membrane sensitive to a specific ion. To use, the tubes are filled with an appropriate electrolyte, inserted into the reusable

holder and connected to a pH meter. WPI says the tips can be replaced in less than a minute, to measure calcium, potassium, tetramethyl ammonium, sodium and chloride ions in solution. The Kwik reusable holder comes with a BNC cable and an electrolyte filling syringe for \$50 (US), and a choice of one of the five Tip electrode kits costs \$50 (US). WPI recommends its MERE 1 reference electrode for use with the Kwik-Tip system.

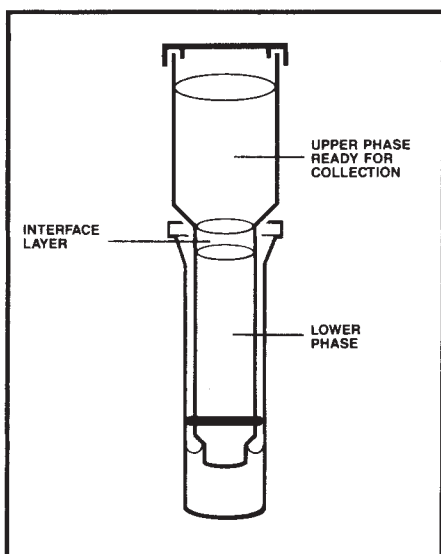
Bulk packs of **disposable pipettes** are available from Sterilin (*Reader Service No. 107*). They are manufactured from polystyrene, and are supplied plugged and sterile, in open-ended or tipped versions. The tipped varieties are made from a single piece of polystyrene, and do not have fluid-trapping shoulders near the tip. The open-ended versions are suitable for fast delivery of volumes, or for pipetting



Sterilin's rainbow of disposable pipettes.

viscous fluids. The pipettes are available with or without suction adapters, and are colour-coded according to size. They come in 1, 2 and 5 ml sizes, in packs of ten.

Xydex has a disposable product for **liquid extractions** — the PhenX (*Reader Service No. 108*). The PhenX can be used for phenol extractions of DNA from lysates, plasmid preps and agarose gels, and for the purification of lipids from serum, plasma samples or cell membranes. The \$42 (US) PhenX kit consists of 48 sets of three sample tubes and a special plunger. After mixing and centrifugation of the extraction phases, the plunger is pressed into the sample tube, forcing the aqueous layer into the plunger lumen and upper compartment. The separated aqueous phase can then be poured or pipetted into another container, leaving the interface layer and organic phases

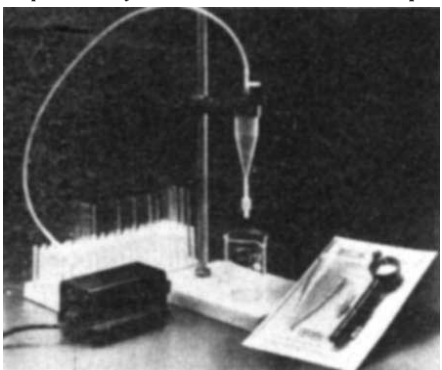


For phenol extractions, and more.

behind. Xydex says the PhenX performs microextractions of 3 ml samples with recovery rates of over 40 per cent.

Elkay is offering free **disposable cuvette** sample packs to laboratories willing to evaluate the six different cuvettes manufactured by the company (*Reader Service No. 109*). The cuvettes are moulded in UV inhibitorless polystyrene or acrylic copolymer, and are capable of averaging 70 per cent light transmittance at 340 nm, says Elkay. The cuvettes are supplied in particle-free shrink-wrapped reusable trays. The six cuvettes in the pack include round and square types, made from polystyrene and acrylic.

A disposable **solid-phase extraction system** for sample preparation is available from Waters, a division of Millipore (*Reader Service No. 110*). The \$5.95 (US) Sep-Pak system kit contains sample



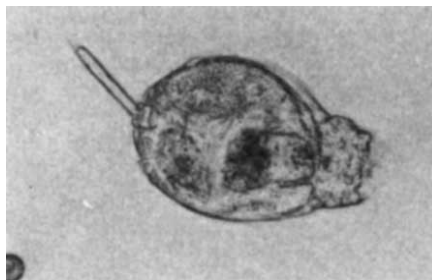
Waters is now offering free samples.

reservoirs, reservoir supports, C₁₈ cartridges and an air pressure pump. The C₁₈ cartridges provide sample preparation in a variety of chemistries, and connect to the bottom fitting of the sample reservoir. Solution volumes ranging from 0.5 ml to 30 ml can be extracted and filtered using the Sep-Pak. Waters is offering free samples of the Sep-Pak system to researchers who would like to try it out.

Human genetics

Applied Biosystems will be exhibiting its line of instruments for the synthesis, isolation, purification and sequencing of nucleic acids and peptides in booths 233 and 235. Sure to be on display is the model 430A **peptide synthesizer** (*Reader Service No. 111*). Fmoc chemistry is now available for the \$79,500–87,500 (US) system, complete with synthesis protocols and a complete range of "plug-in" chemicals. The system can be adapted for t-Boc chemistry — recently introduced with an alternative to HF — by simply changing the floppy disk and loading the relevant chemicals. The model 430A synthesizer can be used with Applied Biosystems' model 151A peptide purification system for a complete system for peptide synthesis, from chain assembly to cleavage and final purification.

Nikon has a new $\times 60$ **phase contrast dry objective lens** that extends the magnification range of the Diaphot binocular laboratory microscope to $\times 900$, using an $\times 15$ eyepiece (*Reader Service No. 112*). The



Shot of a rotifer, 300 μm in length, taken with Nikon's Diaphot with the new $\times 60$ lens.

lens extends the Diaphot's resolution over a long working distance of up to 3.07 mm. Cover glass thickness compensation is included in the Diaphot microscope, and there is no need for a special turret condenser. Nikon also has two new phase plan achromatic objectives, with magnifications of $\times 20$ and $\times 40$. The $\times 20$ has a working distance of up to 6.0 mm, and the $\times 40$'s working distance ranges from 5.08 mm to 6.48 mm, according to cover glass thickness. Nikon will be displaying a full range of its microscopes and objectives in booths 135 and 137.

For the early detection and monitoring of **prostatic cancer**, Hybritech has the Tandem-R PSA immunoradiometric assay (*Reader Service No. 113*). The assay measures serum prostate specific antigen (PSA) — shown to be a more sensitive marker for prostatic cancer than prostatic acid phosphatase (PAP) — quantitatively. Hybritech says the assay achieves a minimal detectable PSA concentration of less than 0.15 ng ml⁻¹, and is linear to 100 ng ml⁻¹. Clinical trials at several major cancer research centres, including Johns Hopkins Hospital and the Memorial Sloan-Ketter-

ing Institute, demonstrated that 95.7 per cent of patients had continually decreasing levels of PSA after successful therapy, as shown by the Tandem-R PSA assay. The assay also indicated tumour recurrence: 94.6 per cent of patients exhibited PSA elevations prior to and during the onset of progressive or recurrent prostate cancer. The Tandem-R PSA assays cost \$590 (US) for an antibody and a substrate set for 100 determinations. Hybritech will be in booth 230.

Hoefer Scientific has a bulletin that describes a **technique for DNA microassays** (*Reader Service No. 114*). Small amounts of DNA are difficult to quantify especially in the presence of RNA or protein. Hoefer's Technical Bulletin Number 119 explains how to determine nanogram to milligram amounts of DNA in dilute samples, using the fluorochrome Hoechst 33258. The technique can be used with Hoefer's IKO 100 DNA minifluorometer, or other fluorometer. Hoefer will be exhibiting its line of instruments for gel electrophoresis in booth 236.

To help valuable DNA restriction and modifying enzymes keep their cool on the way from the freezer to the lab bench, Stratagene has a **portable enzyme storage box** called the StrataCooler (*Reader Service No. 115*). The \$165 (US) StrataCooler holds up to 24 standard microcentrifuge tubes, and keeps them at a temperature of -20°C . The box is constructed of durable acrylic filled with a special freezing agent that maintains the temperature inside at less than -15°C for over two hours on the bench, according to Stratagene. Stratagene will be displaying its wares in booth 127. □

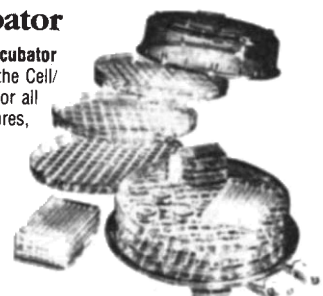
These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further information, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENT

Versatile Tissue Culture Incubator

Modular Incubator Chamber (the Cell/Sphere™) for all tissue cultures, aerobic or anaerobic.

Low Cost



Eliminates evaporation loss and contamination. Easy to use — flush with gas mixture, seal, place at appropriate temperature. Stackable, see-through. Quantity discounts. billups-rothenberg, inc., pob 977, Del Mar, CA USA 92014. Call 619-755-3309 or contact your Flow Laboratories representative.

Reader Service No.174

Less unemployment, or skills shortage

Richard Pearson

Population changes in the United Kingdom during the next decade may help solve one problem but will create others which are difficult to predict.

WE are now entering an era of sharp demographic change. Across most of the non-catholic Western world (though not in Japan), the number of 16–18-year-olds in the population will decline rapidly over the next decade as a result of the slump in births in the 1960s. In the United States there is a 23% decline between 1908 and 1994, while in Germany the decline is a massive 44%. In the United Kingdom the downturn will be 26% over this period, with most of the decline during the early 1990s (*Nature* 326, 528; 1987).

Is this decline going to be the answer to the United Kingdom's long-term problem of high youth unemployment? In areas of high unemployment perhaps, but the consequences for some employers, course organizers and training organizations will be profound. Already the health service cannot recruit enough school leavers into nursing, while many of the government's Youth Training Schemes, set up to improve the vocational skills of young people, have empty places because better paid full-time jobs attract leavers in many areas of the South East. These problems are being experienced when the number of school leavers is near its peak and the decline has barely begun.

Major differences in the birth rates between social classes — with differing occupational aspirations and qualifications — are an added complication. Because the decline in the birth rate among the higher social classes has been less severe (Table 1), the decline in better qualified leavers will also be less severe. So the reduction in demand for places in higher education will be less than crude numbers suggest. Unfortunately, while forecasts of school leavers with differing qualification profiles are available up to the early 1990s (Table 2), changes in the exam structure make longer term forecasts extremely speculative.

What then are the most important implications of these trends, and which sectors need to adopt new policies to cope? Clearly, at the top end of the qualification profile competition for recruits is going to become ever more fierce.

The banks, for example, who take many thousands of well qualified school leavers each year, will be in direct competition with technology courses in higher education for the school leaver with two A

Table 1 Projections of number of 18-year-olds in the United Kingdom by social class (thousands)

Year	Social class						Total
	I	II	IIIN	IIIM	IV	V	
1980	56	186	93	329	165	52	881
1985	66	202	93	343	145	46	893
1990	70	213	85	311	96	30	801
1995	63	183	62	215	72	19	613
2000	61	213	67	242	73	25	681

Key: I, professional; II, intermediate; IIIN, skilled non-manual; IIIM, skilled manual; IV, semi-skilled; V, unskilled. Source: AUT

levels. A factor which may make things worse for organizers of technology courses is a possible replication in the United Kingdom of the trend in the United States where students are turning away from engineering and technology (*Nature* 329, 90; 1987).

Further down the qualification profile the health service is going to face a major challenge. Even now the health service cannot recruit enough nurses, and with the potential pool of qualified candidates shrinking the prospects are poor.

What can be done to mitigate the unwanted side effects of the post-baby-boom era? A first step is for an organization to be aware of its dependence on school leavers, and numbers are not the sole guide. An employer in the South East

example, are five O levels really necessary for a given job? It may be that entry standards have been raised in response to the easier recruitment environment of the past few years. Qualifications are also often seen as the only screen for intelligence and competence; the Youth Training Scheme has now demonstrated that alternative forms of experience can be just as relevant. Likewise is a physics or mathematics A level really necessary, what about the combined science qualifications or an aptitude or numeracy test?

Recruitment can also be redirected to other groups; one new source of skills that has been opened up by the recession and the increasing rate of change in the labour market is that of the mature (re)entrant or individual seeking retraining. The potential here is enormous and these people can often bring new insights and additional skills to a job. The obvious source here is women returning after child rearing. Finally recruitment levels may not need to remain at traditionally high levels. A reassessment of job content, induction training, selection, job previewing and type of entrant can all contribute to lowering wastage during training and early careers, thus

reducing the need for replacements as well as improving job satisfaction for the individual.

The demographic downturn is going to provide both opportunities and problems. While reductions in youth unemployment will of course be welcome, many recruiters for both jobs and training courses will have to think long and hard about innovation and overturning tradition if they are to continue to meet their employment needs during the next decade and beyond. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.

Table 2 School leavers available for employment (1979/80–1992/93)

Examination achievement	Percentage decline 1979/80–1992/93
A levels	– 21 %
5 or more 'good' O levels*	– 34 %
1–4 'good' O levels	– 40 %
Other O level/CSE	– 35 %
No graded results	– 58 %
Total	– 40 %

* 'Good' O levels are O level passes graded A to C and CSE grade 1. Forecasts for 1987/8 and beyond are complicated by the existence of GCSE. Source: DES/IMS.

seeking five trainee technicians with science O levels may face far greater difficulty than a high quality Youth Training Scheme in the North West seeking 50 trainees. It is line managers and department heads who also have to be aware of these changes and not try to maintain rigid, historic recruitment criteria in the face of this changed external environment. Innovation in recruitment, training and employment will be needed and has to be a joint responsibility between them and the recruiters.

Next is the need to review the reasons that lay behind the decision to recruit a specific category of school leaver. For